Analysis of Climate Information to Support US Fish and Wildlife Service Species Status Assessments: An Eye-Tracking Study

by

Haven J. Cashwell

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
May 1, 2021

Approved by,

Dr. Karen S. McNeal, Chair, Professor of Geosciences, Auburn University
Dr. Chandana Mitra, Associate Professor of Geosciences, Auburn University
Dr. Christopher G. Burton, Assistant Professor of Geography, University of Connecticut

ABSTRACT

Species Status Assessments (SSAs) are required for endangered species by the United State Fish and Wildlife Service (USFWS). These documents need to include climate information since climate will be negatively impacting species in the future. To aid in including climate information a Decision Support System (DSS) entitled Climate Analysis and Visualization for the Assessment of Species Status (CAnVAS) was developed. The goal of the current study is to fill the gap in assessing co-production and evaluating SSAs. This research analyzes the differences in usability between two versions of a DSS aimed at benefitting the USFWS through eye-tracking and subsequent interviews with novice users. It was found that graphically displaying temporal climate information through box and whisker plots and spatially through a sequential color ramp from white to purple was more effective at communicating climate information on endangered species.

DEDICATION

Dedicated to two of my biggest cheerleaders who always believed in me and encouraged me,
Dennis Dugan & Sandra Freeman Dugan, whose lives were taken far too soon

ACKNOWLEDGEMENTS

I am so thankful for those who have supported me throughout this journey. First, I would like to thank my parents, David and Alice, and my little sister, Nicole. The three of you have been amazing at helping me through my schooling and I could not be more thankful for the constant encouragement. My friends from high school and undergraduate have always been a phone call or text away when needed. I have developed such great friends through my time here at Auburn both within the department and the Auburn Geocognition Lab. Ally, Elijah, Lindsay, Nick, Steph, and Tyler, thank you for being a second family to me while I have been pursuing this degree.

This project was funded by the Southeast Climate Adaptation Science Center, United States Geological Survey Subaward # 2020-0455-01. Dr. Dello, Dr. Boyles, and Corey Davis were instrumental in discussing study design and procedure. I specifically would like to thank to Dr. Maudlin for helping with eye-tracking questions when they arose. I would like to extend gratitude to the research participants in this study. Thank you to my committee members, Dr. Mitra and Dr. Burton. Lastly, thank you to my advisor, mentor, and committee chair, Dr. McNeal, for your dedication to my success; it really has meant the world to me.

TABLE OF CONTENTS

Abstract ... 2

Dedication ... 3

Acknowledgements ... 4

List of Tables ... 7

List of Figures ... 8

List of Abbreviations ... 9

Introduction ... 10

Background ... 11

Objectives ... 14

Materials and Methods ... 15

Species Status Assessments ... 15

Development of CAnVAS 1.0 ... 16

Eye-Tracking and Interviews ... 17

Results ... 22

Species Status Assessments ... 22

Eye-Tracking: Bulk Analysis ... 24

Eye-Tracking: TFF and TFD ... 28

Eye-Tracking: Commonly Missed Questions ... 33

Usability: Efficiency, Effectiveness, Satisfaction ... 38

Discussion ... 43

Species Status Assessments ... 43

Eye-Tracking ... 44

Usability: Efficiency, Effectiveness, Satisfaction 45

Conclusion 48

Future Research 49

References 51

Appendix A 55

Appendix B 56

Appendix C 57

Appendix D 58

Appendix E 60

LIST OF TABLES

Table 1: Codes and Excerpts from SSAs 23

Table 2: Bulk Analysis 27

Table 3: Analysis of TFF 28

Table 4: Analysis of TFD 29

Table 5: Analysis of TFF for Commonly Missed Questions 34

Table 6: Analysis of TVD for Commonly Missed Questions 35

Table 7: Analysis of TFMC 39

Table 8: Effectiveness Analysis 40

Table 9: Codes and Excerpts from Undergraduate Interviews 41

LIST OF FIGURES

Figure 1: The two versions of CAnVAS 18

Figure 2: Examples of AOIs from Task 3 26

Figure 3: Heatmaps for Task 3.3 31

Figure 4: Gaze plots for Task 3.3 32

Figure 5: Average TVD for Task 3.3 33

Figure 6: Heatmaps for Task 1.3 36

Figure 7: Heatmaps for Task 2.3 37

Figure 8: Average TVD for Task 1.3 and 2.3 38

Figure 9: CAnVAS 2.0 48

Figure 10: Heatmaps for Task 2.2 60

Figure 11: Heatmaps for Task 2.4 61

Figure 12: Heatmaps for Task 3.2 62

Figure 13: Heatmaps for Task 3.4 63

Figure 14: Heatmaps for Task 3.5 64

LIST OF ABBREVIATIONS

AOI – Area of Interest

CAnVAS – Climate Analysis and Visualization for the Assessment of Species Status

DSS – Decision Support System

ESA – Endangered Species Act

IRB – Institutional Review Board

SCONC – State Climate Office of North Carolina

SSA – Species Status Assessment

TFD – Total Fixation Duration

TFF – Time to First Fixation

TFMC – Time to First Mouse Click

TVD – Total Visit Duration

USFWS – United States Fish and Wildlife Service

INTRODUCTION

Climate change affects a copious amount of people whether they realize it or not. Right now, climate change is typically seen as a problem that will affect the public some time far into the future (Corner et al. 2018). To many members of the public, climate change is considered to be uncertain as well as very ambiguous (Nisbet 2009). As such, communication is this field has become urgent, but it is one of the most challenging forms of communication in the sciences (Somerville & Hassol 2011). One way to communicate climate change is through images and graphs but visuals involving climate change need to be carefully considered and implemented in order to be most effective with their intended audiences (Corner et al. 2018; Atkins & McNeal 2018; Harold et al. 2016).

One target audience that requires accessible climate information that is effectively communicated are managers and policy makers as they must make decisions about the impacts of climate change on the systems they manage. Decision Support Systems (DSS) are a tool that can be used to bridge the information gap for managers and policy makers and can specifically aid with communication about climate change. “Decision support provides a link between decision making, scientific information, and analytical tools” (Pyke et al. 2007). DSSs can take many forms which can be documents, software tools, or models (Pyke et al. 2007). Some examples of well known DSSs are the NOAA Climate Explorer (Climate Explorer 2016) and the USGS GeoData Portal (Blodgett et al. 2011). DSSs have information that will help users, specifically stakeholders in a particular field, with understanding facts and making critical decisions. Climate change has been, and is going to continue being, at the forefront of concern for decision makers, especially decision makers involved with protecting endangered species.

Climate change has already played a huge role in effecting different species (USGCRP 2018; Hansen et al. 2006). Species managers must understand what changes are going to effect species in their specified area in order to be proactive and help the species through effective management decisions. DSS tools can assist species managers in their decision making by providing the most up to date climate science information while integrating the impacts and behaviors of key regional species. A framework specifically developed for the USFWS, in the form of a DSS, is needed to assist proper management of critical species by linking decision making, scientific information, and analytical tools (Pyke et al. 2007).

This research involves the development and evaluation of a DSS called Climate Analysis and Visualization for the Assessment of Species Status (CAnVAS) as well as evaluation of Species Status Assessments (SSA). The graphics in CAnVAS are directed towards climate change, focusing on its impacts that the United States Fish and Wildlife Service (USFWS) stakeholders must understand, and how climate change affects certain species. It is within this context that this research contains an eye-tracking component to gain valuable insights into how users interact with climate-related information. The eye-tracking portion of this study contributes to the improvement of decision-making involving climate change and species within the USFWS by using the CAnVAS framework. The evaluation of SSAs portion of this research contributes to understanding the climate information that is being included in SSAs and how this can ultimately be improved to add knowledge about how climate change is related to endangered species.

BACKGROUND

Okan et al. (2016), explained that graphical displays are effective DSSs if they aid in comprehension of information. Graphical displays can be bar graphs, scatter plots, etc.. Previous research has tested differences between comprehension of bar and line graphs through a mixed design as well as research on evaluating graph comprehension and pattern recognition through eye-tracking (Shah & Freedman 2011; Carpenter & Shah 1998). It will be beneficial to analyze other types of graphical displays of information to test different graphical displays (i.e. faded bar version box and whisker plot) through eye-tracking and a mixed methods approach to understand the differences in comprehension of quantitative climate information that will be displayed on DSSs.

Color ramps are another important part of the CAnVAS DSS that is used in this study and thus must be evaluated. Choosing the correct representation for specific climate information on CAnVAS was deemed a priority to ensure an evaluation between groups. Previous research has identified that "Rainbow" color ramps are not substantial enough for displaying information thus were ruled out of consideration for color schemes for CAnVAS (Turton et al. 2017; Moreland 2009). When choosing colors for interpretation, it is essential to consider the dimension of color as well as the background that the color ramp is being displayed on (Schloss et al. 2018). Using an accurate, but informative color ramp will be helpful for displaying climate information on CAnVAS.

The USFWS has specific regulations they must abide by under the Endangered Species Act (ESA). Under the ESA, the USFWS is required to evaluate the status of at-risk plants and animals in the United States. The way the USFWS conducts this evaluation is through an extensive document called a SSA. A SSA is prepared for at-risk species to help inform a range of management decisions under the ESA. SSAs contain information about the species habitat, what the species is going to experience in the future, its current condition, etc. Having knowledge about long-term climate change impacts on species is important for species managers, especially those that prepare SSAs (Hulme 2005).

In order to help USFWS biologists incorporate climate information into SSAs, the State Climate Office of North Carolina (SCONC) developed the CAnVAS framework, which acts as a DSS, in the form of a website. This DSS was based on a previous design created by the SCONC which was called PINEMAP (Alridge et al. 2016; Davis et al., 2020). Research conducted on PINEMAP by Davis et al. (2020) found that a three-map layout and specific location-based information was efficient at communicating climate information. The climate information provided in CAnVAS includes how climate factors and thresholds that most affect species vary year to year, how they are expected to change in the future, and the uncertainties associated with those changes while using specific location-based information. The SCONC decided which graphics to incorporate within the website and consulted biology experts to determine which species to target on CAnVAS for this first iteration of research.

The SCONC designed CAnVAS due to the "…prompted transitions from scientist to stakeholder-driven or collaborative approaches to climate science" (VanderMolen et al. 2020), or co-production. Co-production involves science makers collaborating with science users to create a final product that can be successfully used in decision making to produce usable climate science knowledge (Meadows et al. 2015). Although co-production approaches have been recently used in actionable science, the products that result from these efforts are often not rigorously evaluated for their usability or use with the intended stakeholders (Wall et al. 2017; Durose et al. 2014). By providing information from initial trials of this research, SCONC will improve CAnVAS which builds upon the process of co-production.

OBJECTIVES

The overarching goal of this project is to fill the assessment gap in co-produced research by designing a web-based DSS for species managers and evaluating its usability when applied to the construction of SSAs. The main outcome of co-production is usability, and this outcome should be tested in new ways that align more with its role in improving science and outcomes in the society (Bell et al. 2011; Meadows et al. 2015). The research questions guiding this project include: (1) What types, how much, and with what accuracy is climate information currently included in SSAs in order to understand the effects climate change is having on endangered species in the southeast US? (2) How do novice users engage with different versions of CAnVAS? (3) How can the usability of CAnVAS be improved? These questions were answered through a mixed methods approach though coding and ranking of SSAs as well as eye-tracking and interviews with undergraduates. Research question (1) was answered through thematic coding and ranking of SSAs. Research questions (2) and (3) were answered by eye-tracking novice users engaging with CAnVAS. Eye-tracking allows for correlating "…eye-tracking metrics used and theoretical constructs of interest" and assessing usability (King et al. 2019; Haesnar et al. 2018; Ehmke & Wilson 2007). Eye-tracking measured the usability of CAnVAS by looking at three metrics: efficiency, effectiveness, and satisfaction (Maudlin et al. 2020; Courtney 2019 p. 7; Creager & Gillan 2016). Efficiency was measured by *how long* it takes for a participant to choose an answer based on post-DSS viewing questions provided during the eye-tracking experiment. Effectiveness was measured by participant *performance scores* on post-DSS viewing questions.. Finally, satisfaction with CAnVAS was measured by the overall participant *perceived usefulness* of the DSS through retrospective interviews with participants.

MATERIALS AND METHODS

Species Status Assessments

To inventory the amount and type of climate information included in SSAs, 48 SSAs were downloaded from the USFWS Southeast region website. The requirement for these SSAs to be included on this particular website was that either the species habitat was located in the Southeast or one of the authors worked in the Southeastern regional office for the USFWS. Consequently, some of the SSAs that were analyzed for this study included species that were located in other areas besides the Southeast. For instance, an example of a SSA not in this region but included on the website would be the assessment focused on the Freemont County Rockcress located in Wyoming.

A coding scheme and ranking rubric for SSAs had to be developed which are both included in Appendix A and Appendix B. The coding scheme and ranking rubric were created after reading through three SSAs and noting what climate information was both included and missing in the assessments. Coding allows for the data to be broken up into smaller parts and find themes amongst those parts which can be labeled (Schwandt 2015). The coding scheme included both parent codes and child codes. Parent codes were generalized topics included in SSAs and the child codes were more specific codes that were categorized under the parent codes. The ranking rubric used the same parent codes for categories and then scored those categories for accuracy, quality and depth, and relevancy. This coding scheme was reviewed by external experts in climate science, co-production, and geocognition before use and the ranking rubric was developed from the coding scheme. These SSAs were uploaded into the thematic coding program Dedoose for further analysis. In order to obtain a confirmation of agreement, thematic co-coding was completed for 10% of the codes which was conducted by a co-member of the

Auburn Geocognition Lab. Co-ranking was done for 10% of the SSAs which was also conducted by a member of the Auburn Geocognition Lab.

Development of CAnVAS 1.0

For the development of the CAnVAS DSS, the SCONC met with a committee of stakeholders associated with the USFWS to assess needs for this DSS, an important step in the development process (McNie 2012). Once a preliminary DSS framework was developed, the representatives of the SCONC met with researchers at Auburn University to discuss the first iteration of CAnVAS and the eye-tracking experimental approach to be used in the research study. The first iteration of CAnVAS was still image layout rather than a "live" navigational website and had representation for just one endangered species – the gopher tortoise. This decision was made in part so that the evaluation could provide important feedback on visual and graphical features before aspects of CAnVAS became difficult to change in a more dynamic website format. Similarly to Hegarty et al. (2010), an A/B study design was developed where two versions of CAnVAS were made to allow for testing of specific design features. Two visual design differences were examined between version A and version B – how data was displayed in the climate projection portion (i.e. faded bar/box plot) and the climate snapshot color scheme (i.e. white to purple/blue to gray). These design features were chosen for manipulation as previous research points to testing differences between graphical designs within DSSs to ensure effective communication and determining the best sequential color ramps to use to depict climate information (Schloss et al. 2018; Turton et al. 2017; Moreland 2009; Okan et al. 2016; Shah & Freedman 2011; Moreland 2009; Carpenter & Shah 1998).

Eye-tracking and Interviews

Eye-tracking and interviews of undergraduates is an established method for improving and designing web-based applications in order to effectively communicate scientific information (Davis et al. 2020; Mason et al. 2012). In this study, eye-tracking and interviews contribute to improving CAnVAS, before implementation to stakeholders in the USFWS, in order to ensure that it is effective for communicating on climate change impacts to endangered species. For this portion of the study, eye-tracking and interviews align with the sequential explanatory strategy indicative of a mixed methods design. This strategy involves "…the collection and analysis of quantitative data in a first phase of research followed by the collection of and analysis of qualitative data in a second phase that builds up on the results of the initial quantitative results" (Creswell 2009). This design strategy is used to explain the quantitative data, eye-tracking data, by collecting the qualitative data, interview data.

Before recruitment of participants began, four members of the Auburn Geocognition Lab participated in a pilot study to refine details of the study procedure as well as identify if any obvious changes needed to be made to CAnVAS. Only one change was made to the box and whisker version of CAnVAS which was changing the wording of the labels on the climate snapshots at the bottom to match the wording of the faded bar version of CAnVAS. The data from the pilot study was not used for analysis and was strictly used for altering the study procedure and changing wording on CAnVAS. The two final versions of CAnVAS that were used for this study are shown below in Figure 1. We will use the terminology "box plot" to indicate Figure 1 (top) and "faded bar" to indicate Figure 1 (bottom) designs from herein.

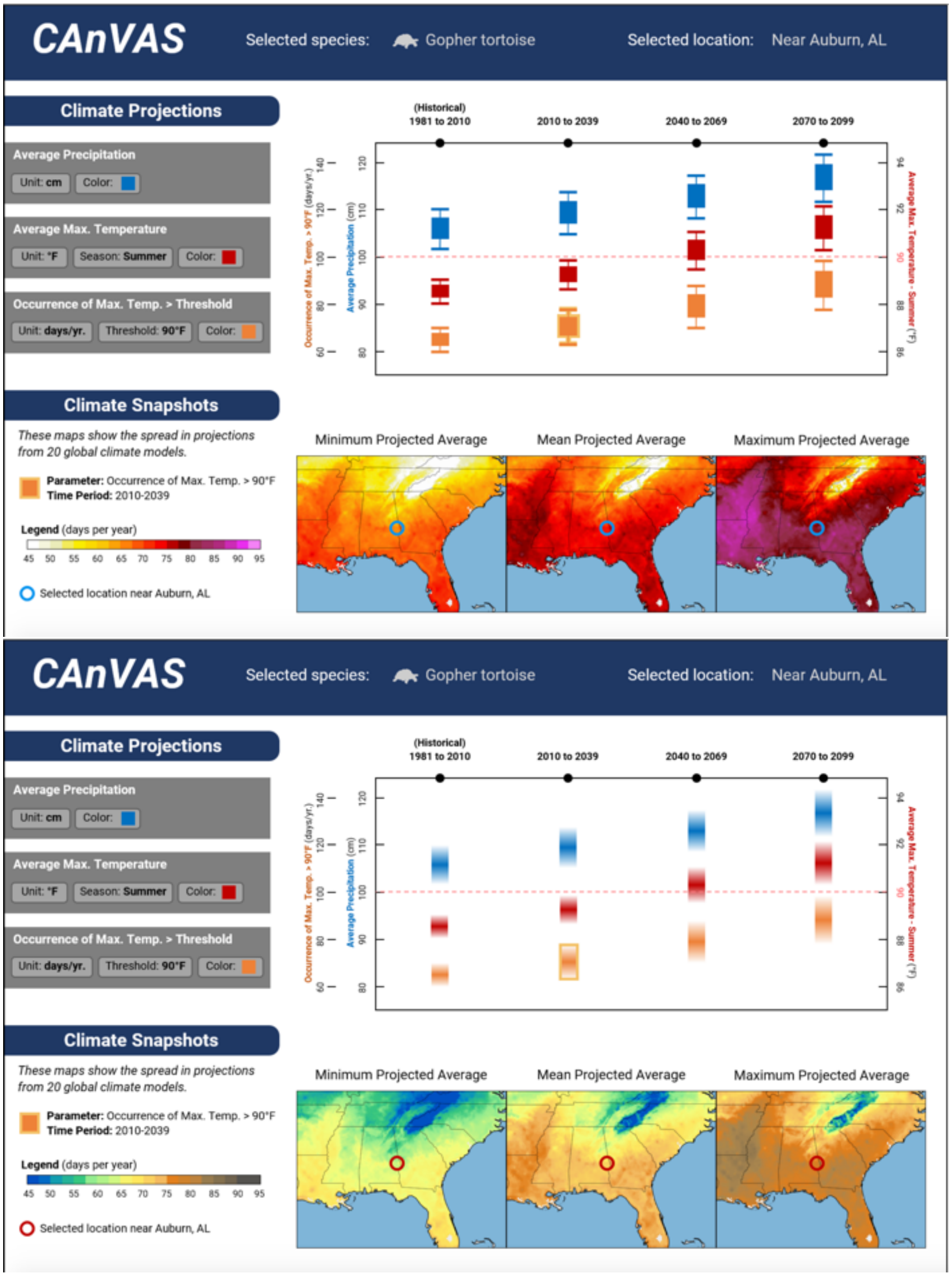

Figure 1: The two versions of CAnVAS to be tested with undergraduates.

CAnVAS depicts how the gopher tortoise species is going to be impacted by different variables (average maximum temperature, average precipitation, occurrence of maximum temperature) of climate change over time. The location for the gopher tortoise analysis was determined, before testing, as “Auburn, AL” so that participants would be familiar with the location when they were asked to look at Auburn during Task 3. Occurrence of maximum temperature data from 2010-2039 is displayed in the climate snapshot, bottom portion of CAnVAS. Both portions of CAnVAS have associated legends on the left-hand side of their associated data. The main differences in the two versions of CAnVAS were the way the data was displayed in the climate projections (faded bar/box plot) and the two different color schemes used in the representation of the climate snapshots. Everything else in the study (axis labels, legends, questions, etc.), remained the same between groups.

Undergraduate participants were recruited for this study through emails, classroom visits over Zoom, and flyers around campus. Emails and classroom visits over Zoom were done in introductory science classes (i.e. Principles of Biology, Dynamic Earth, etc.) across the entire university. Undergraduates were chosen as a proxy for novice users for this research. The goal was to recruit 30 participants for each study group for statistical robustness. Typically, eye-tracking studies tend to have smaller sample sizes, so having a goal of at least 30 participants per group for this study was deemed as an appropriate goal for statistical analysis (Jacob & Karn 2003). Once the initial contact was made with participants that indicated they were interested in the study, the participants were sent an email with the Institutional Review Board (IRB) information letter and a link to sign up for a time slot. Overall, 64 students sent an email saying they were interested in the study and 39 signed up to participate.

Participants came into the Auburn Geocognition Lab in the Sciences Center Classroom Building on their designated time and day. Verbal confirmation was received on whether or not participants read the IRB information letter before being allowed to participate. Then, participants sat at the desktop computer in the lab at approximately 65 cm from the eye-tracker and were calibrated with the TX-300 Tobii eye-tracker allowing for precise measurements of the eye-tracker. The eye-tracker is mounted at the bottom of the desktop computer. The calibration process required that the participant follow red dots across the screen and calibrations of >70% weighted gaze samples were allowed to continue to the main study. Data was collected within the Tobii Studio 3.4.8 software (Tobii, Stockholm, Sweden). It should be noted that for human subjects research, all COVID-19 guidelines instituted by the CDC at the particular time the study was conducted were followed.

Participants were randomly assigned to the two different versions of CAnVAS (i.e. box plot or faded bar). Then they viewed the free exploration components of CAnVAS for 1 minute and 15 seconds to allow participants to get acquainted to CAnVAS before answering questions. There were two free explorations located before Task 1.1 and Task 3.1. Participants answered 15 questions about CAnVAS in order to test their usability of the DSS. All of the questions included six different answer choices. Four of those answer choices could have been possible answers to the questions and the other two were "I am not sure because I did not understand how to solve the question" and "I am not sure because I was unable to interpret the data." This allowed for participants to pick an answer choice based on their confidence, rather than choosing a particular answer because they did not have another option or were not totally certain about the answer. There was no time limit on how long a participant could spend answering a question.

Once participants were done with the eye-tracking portion of the study, approximately 15-minute long retrospective interviews were conducted, and audio recorded with participants. During these interviews, participants viewed their own gaze replay to understand exactly what their eye movements looked like. A gaze replay is "…video replay of the task with eye-movements collected during the tasks superimposed over the stimulus" (Bojko 2013). This follows the retrospective verbal protocol that includes playing some type of visual (gaze replay) and have the participant verbally recall what they were thinking throughout the task (Cho et al. 2019; Bojko 2013). Interview questions that were asked during this section of the research are included in Appendix C. Once all interviews were complete, the audio was removed of random identifiers and transcribed by a transcription service. Then, transcripts were entered into Dedoose and the interviews were thematically coded. The coding scheme used for the interviews is included in Appendix D.

A main limitation of this study was conducting eye-tracking on a still image version of CAnVAS versus having CAnVAS as a navigational website. However, it was necessary that we evaluate the still image, first iteration of CAnVAS, in order to get proper results to ultimately benefit a navigational website to be developed in the future. Also, it would be more difficult to make changes to a "live" website rather than a still image. Another limitation of this design was the length of time which was required in order to obtain comprehensive results. However, to offset the length of time that it took participants, a monetary compensation, in the form of a gift card, was given to participants after the completion of their portion of the research. Recruitment for this study was a limitation due to classes being mostly online during this semester and people's comfortability with doing in person research prevented access to recruit some participants. Participants in this research were not the target audience of the CAnVAS DSS, but it was not a requirement to have the target audience for the first iteration of evaluating CAnVAS. Future research will include eye-tracking with the USFWS stakeholders which will then have an evaluation of the CAnVAS DSS with the target audience. Also, the Tobii Studio software did not record free explorations for some of the participants, thus impacting some eye-tracking data on

Task 1.1 and 3.1 due to those tasks being right after the free explorations. In some analysis, data was normalized to Time to First Mouse Click (TFMC) in the answer Area of Interest (AOI), which is a limitation in that participants could have selected an answer and then reconsidered if that was the correct answer or not before continuing altering the time spent on the web page.

RESULTS

Species Status Assessments

The SSAs varied in length from the shortest in length being 24 pages and the longest in length being 233 pages. Percentages of pages pertaining to a heading of climate change ranged from a minimum of 0% to maximum of 15% of the total reports. If the SSA received 0%, this means that there was no heading with the words “climate change” explicitly mentioned. The average percent of pages of climate information included in a heading of climate change across all SSAs that were pulled from the USFWS Southeastern website was 2.91%.

The coding scheme allowed for an in depth analysis of SSA data. The top three codes that were used the most were: climate change (n = 733), urbanization (n = 662), and increase in temperatures (n = 509). The bottom three codes that were used the least were: uncertainty (n = 5), climate adaptation/resiliency (n = 7), and anthropogenic emissions (n = 19). Co-coding was completed on 10% of codes and resulted in a good agreement (*Cohen’s Kappa* = 0.79). Table 1 below shows the parent codes used in the coding process of SSAs along with an excerpt showcasing how the code would have been represented in the SSA.

Table 1

Codes from Analysis of SSAs with Accompanying Excerpts

Code	Excerpt from SSAs
Climate Change	"Climate change is also expected to change the timing and quantity of stream flows, and anthropogenic use is likely to exacerbate these effects."
Extreme Weather	"Stream velocity is not static over time, and variations may be attributed to seasonal changes (with higher flows in winter/spring and lower flows in summer/fall), extreme weather events (e.g., drought or floods), and/or anthropogenic influence (e.g., flow regulation via impoundments)"
Impacts to Biology	"Cayo Diablo and USVI Cay are most at risk, while there is a moderate risk of sea level rise and storm surge impacts to VI boas and habitat on Cayo Ratones, and low risk at Culebra, Río Grande, and St. Thomas."
Increase in GHGs	"Although significant discussion is aimed at global emissions reduction, emissions continue to escalate and presently there is no clear socioeconomic driver to depart from a carbon-based energy infrastructure."
Increase in Temperatures	"Most of the warming occurred in the past 35 years, with the five warmest years on record since 2010. The oceans have absorbed much of this increased heat, with the top 700 meters (about 2,300 feet) of ocean showing warming of more than 0.4 degrees Fahrenheit since 1969."
Precipitation Pattern Changes	"Changes and shifts in seasonal patterns of precipitation and runoff will alter the hydrology of stream systems, affecting species composition and ecosystem productivity."
Solutions	"Reducing the likelihood of significant impacts will largely depend on human activities that reduce other sources of ecosystem stress to ultimately enhance adaptive capacity; these include maintaining riparian forests and forested wetlands, reducing nutrient loading, restoring damaged ecosystems, and minimizing groundwater and surface water withdrawal."

With the ranking of SSAs, three categories were evaluated: accuracy, quality and depth, and relevancy. In the ranking process, the lowest score given was a 1 and the highest was a 5. If information for that category was not in the SSA, then that category received a non-applicable (N/A) in the place of a value. To determine a representation of the three categories, averages of the scores from the rankings were taken across the three categories that were evaluated.

Accuracy (*M* = 2.79) achieved the highest average, followed by relevancy (*M* = 2.63), and then finally quality and depth (*M* = 2.39). 37 out of 48 SSAs were missing content related to solutions for these endangered species and climate change. There 10 out of 48 SSAs that did not contain information for 3 or more of the code categories that were being ranked. Co-ranking was done on 10% of SSAs and received good agreement (*Cohen's Kappa* = 0.61).

Eye-Tracking: Bulk Analysis

Overall, there were 39 participants that were involved in the eye-tracking portion of this research. However, two participants data were discarded because they did not reach the requirement of at least 70% weighted gaze samples during the study. This left the analysis with 37 participants. Out of those 37 participants, the majority of them were 19 years old (*n* = 19) and white (*n* = 30). A slight majority of the participants were female (*n* = 20). Participants were almost evenly spread across the grade classification between Freshman (*n* = 8), Sophomore (*n* = 11), Junior (*n* = 10), and Senior (*n* = 8). Undergraduates with some likely prior knowledge in climate change and species were targeted and therefore were from the College of Engineering (*n* =10), the College of Forestry and Wildlife Sciences (*n* = 7), College of Business (*n* = 6), College of Sciences and Mathematics (*n* = 5), College of Liberal Arts (*n* = 3), College of Agriculture (*n* = 2), College of Human Sciences (*n* = 1), and School of Nursing (*n* = 1). With random assignment, there were 19 participants in the faded bar version and 18 participants in the box plot version of CAnVAS.

After all participants' completion of the eye-tracking and interviews, AOIs were created for each CAnVAS page. AOIs are drawn, through the Tobii Studio software (Tobii, Stockholm, Sweden), on a still image, and are used to quantitatively measure participants' eye movements across the particular section. According to Bojko (2013), AOIs must reflect study objectives which was taken into consideration for every AOI that was drawn. AOIs remained consistent across both versions of the study meaning that AOIs were the same size and in the same location for both versions. AOIs were drawn for 17 web pages since all questions and free explorations were located on a separate page. For some pages, smaller AOIs were drawn within larger AOIs in order to get an accurate representation if participants were looking at the correct data when trying to find the correct answer to the question. An example of AOIs for one question of both the faded bar and box plot versions of CAnVAS are located below in Figure 2. In this figure, an example is shown of an instance where smaller AOIs were drawn inside of larger AOIs.

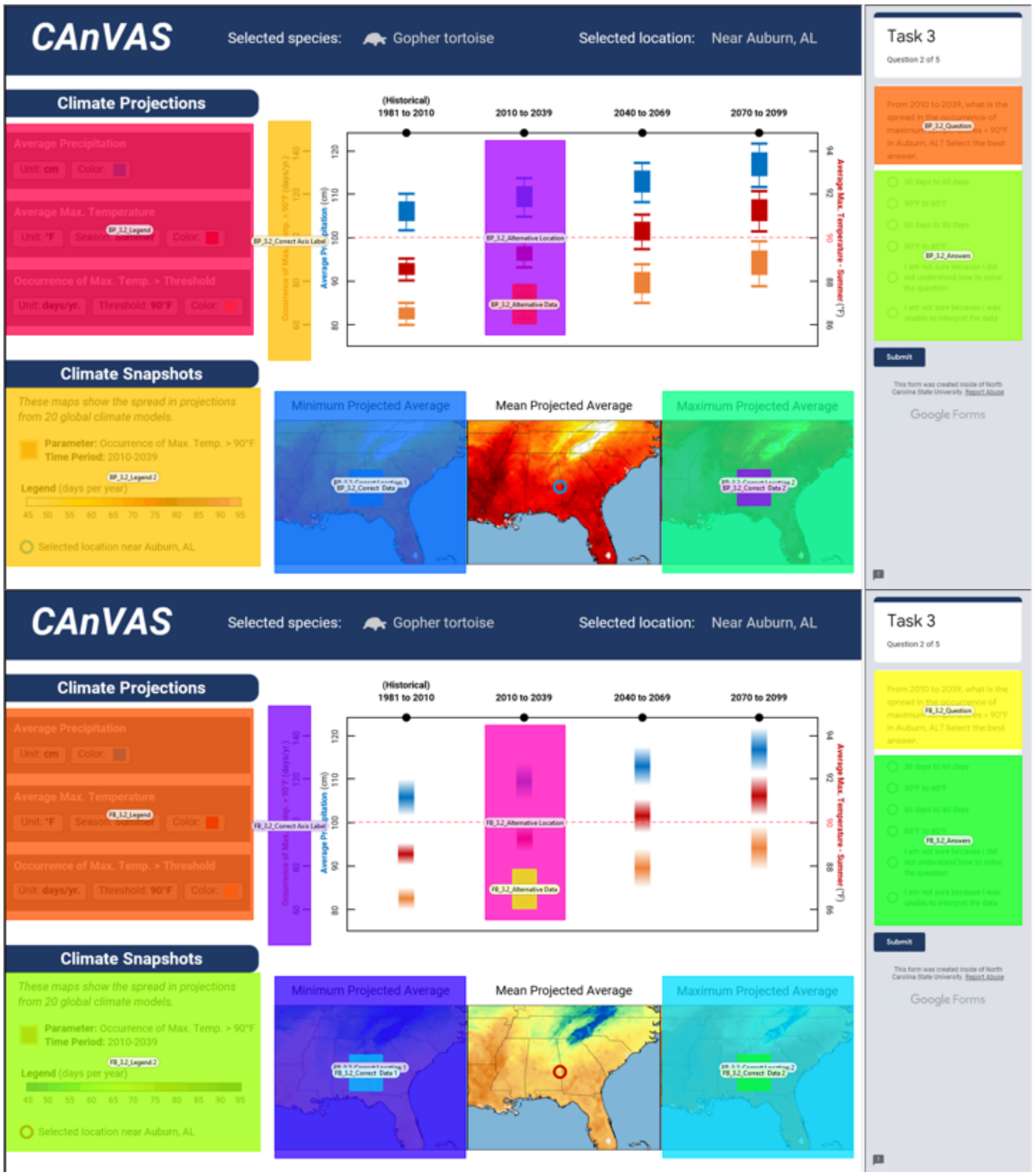

Figure 2: Examples of AOIs from Task 3 of CAnVAS.

AOIs, in this research study, were used to measure the following eye-tracking metrics: Total Fixation Duration (TFD), Time to First Fixation (TFF), Total Visit Duration (TVD), and Time to First Mouse Click (TFMC). TFD, TFF, and TVD were normalized to TFMC. TFMC was used as a proxy for how long a participant spent on the particular question since at the TFMC in the answer AOI, the participant would have selected an answer. Normalizing this data was crucial to get an accurate representation of how long a participant would have spent in the other AOIs that were measured relative to the time spent on the webpage.

To begin looking at the eye-tracking data, a bulk analysis was done for TFD, TFF and TFMC. These eye-tracking metrics were summed, per group, across all tasks. These metrics were not normalized since this was an initial bulk analysis and the goal was to understand if there were any obvious differences between groups before implementing a finer analysis, which is described later. These totals were then analyzed through an independent samples t-test to measure significance and to calculate Cohen's d, a measure of effect size. A negative effect size indicates that the mean for the box plot version of CAnVAS was larger than the mean for the faded bar version of CAnVAS. Table 2 shows the results of the bulk analysis conducted.

Table 2

Bulk Analysis Between Groups Analysis of Eye-Tracking Metrics

	Condition	**Sample Size**	**Mean**	**Std. Dev.**	**Sig.**	**Effect Size**
TFD – Correct Locations	Faded Bar	19	0.96	0.44	.105	0.55
	Box Plot	18	1.20	0.42		
TFF – Correct Locations	Faded Bar	19	6.66	1.74	.987	0.01
	Box Plot	18	6.67	2.02		
TFMC	Faded Bar	19	456.90	173.90	.492	0.23
	Box Plot	18	420.62	141.04		

Eye-Tracking: TFF and TFD

In eye-tracking, TFF is how long it took for a participant, from the start of the question, to fixate on a certain AOI. In this case, TFF was measured on the correct locations AOI. TFF for correct locations helped to understand how long it took participants from both groups to look at the correct spot on CAnVAS. Correct location, in this study, is defined as where the participant would have had to look to retrieve the correct answer to the question. For example, in some instances where participants had to look at a particular piece of data, they would have had to understand to look at the axis label to retrieve the correct answer which would mean that the axis label would be the correct location. TFF for all correct locations were summed up and evaluated through an independent samples t-test in order to test for overall statistical significance. According to Table 2, there was no statistical significance nor large effect size between groups. To have a robust analysis, effect size was looked at for each individual correct location on every task for TFF. Per Table 3 below, there were 5 correct locations that had large (>|0.5|) effect sizes: Task 2.2, Task 3.2 B, Task 3.3, Task 3.4 A, and Task 3.5. Task 2.2, 3.2, 3.4, and 3.5 have heat maps for both versions of CAnVAS located in Appendix E.

Table 3

Analysis of Time to First Fixation on Correct Locations for both versions of Climate Analysis and Visualization for the Assessment of Species Status

Correct Location Task	TFF Avg. – FB	TFF Std. Dev. – FB	TFF Avg. – BP	TFF Std. Dev. – BP	TFF Effect Size
1.1	0.61	0.36	0.46	0.27	0.48
1.2	0.49	0.21	0.45	0.27	0.17
1.3	0.43	0.33	0.44	0.27	0.03
1.4	0.46	0.21	0.45	0.17	0.05
1.5 A	0.55	0.57	0.64	0.53	0.16
1.5 B	0.43	0.47	0.50	0.48	0.13
2.1	0.36	0.23	0.34	0.23	0.12
2.2	0.68	0.33	0.46	0.38	0.64
2.3	0.61	0.46	0.49	0.24	0.35

2.4	0.62	0.20	0.51	0.27	0.47
2.5	0.75	0.47	0.89	0.79	0.22
3.1	0.55	0.48	0.64	0.43	0.21
3.2 A	0.59	0.37	0.62	0.34	0.09
3.2 B	0.93	0.59	0.62	0.37	0.69
3.3	0.64	0.18	0.47	0.26	0.73
3.4 A	0.46	0.35	0.23	0.18	0.82
3.4 B	0.59	0.35	0.64	0.26	0.18
3.5	0.60	0.35	1.2036	1.57732	0.57

Note: For 3 questions, there were 2 correct locations since those questions were about spread and required two different values in the answer.

The next eye-tracking metric that was analyzed was TFD for correct locations on each question. TFD is how long the participant spent fixated in a certain AOI. Participants could fixate on a certain AOI more than once by coming back and forth to that AOI. Analyzing TFD for correct location allowed the researchers to understand how long users focused on the correct location. Once again, the same correct locations were used for this analysis. As previously mentioned, TFD for all correct locations were summed up and evaluated through an independent samples t-test in order to test for overall statistical significance between groups. Per Table 2, there was no statistical significance between groups. However, there was a large effect size for the bulk analysis of TFD for correct locations. To analyze TFD for correct locations further, effect size was calculated for each individual task and the results are located in Table 4. A large effect size was found for correct locations for Task 2.4, Task 3.3, and Task 3.4 B. Task 2.2 and 3.4 have heat maps for both versions of CAnVAS that are included Appendix E.

Table 4

Analysis of Total Fixation Duration on Correct Locations for both versions of Climate Analysis and Visualization for the Assessment of Species Status

Correct Location Task	TFD Avg. – FB	TFD Std. Dev. – FB	TFD Avg. – BP	TFD Std. Dev. – BP	TFD Effect Size
1.1	0.09	0.04	0.11	0.10	0.19
1.2	0.12	0.06	0.13	0.06	0.08
1.3	0.08	0.06	0.11	0.08	0.35

1.4	0.15	0.08	0.13	0.05	0.23
1.5 A	0.08	0.09	0.10	0.07	0.30
1.5 B	0.07	0.06	0.01	0.08	0.08
2.1	0.04	0.02	0.06	0.07	0.35
2.2	0.04	0.04	0.06	0.03	0.43
2.3	0.09	0.07	0.10	0.09	0.04
2.4	0.09	0.04	0.14	0.09	0.71
2.5	0.04	0.02	0.05	0.07	0.30
3.1	0.03	0.02	0.03	0.02	0.24
3.2 A	0.03	0.03	0.02	0.02	0.19
3.2 B	0.03	0.01	0.03	0.04	0.04
3.3	0.05	0.02	0.09	0.07	0.73
3.4 A	0.08	0.07	0.09	0.07	0.15
3.4 B	0.03	0.03	0.05	0.03	0.63
3.5	0.05	0.05	0.03	0.04	0.34

Note: For 3 questions, there were 2 correct locations since those questions were about spread or asked for two different values in the answer.

The one task that had a large effect size with the correct location for both TFF and TFD was Task 3.3. This task asked for the participant to find the mid-range of the model projections for the occurrence of maximum temperature greater than 90 degrees Fahrenheit during the 2010 to 2039 period in Auburn, AL. This required the participant to look at the middle panel in the climate snapshots on the bottom half of CAnVAS and focus on Auburn, AL to see what value correlates with the color in the legend on the left. Participants could have looked at an alternative location in the climate projections to find the answer, but the main goal was to have the participant look at the climate snapshots during all of Task 3. In order to properly understand the differences between the two groups, heatmaps of both versions of CAnVAS are provided below in Figure 3. Also, one participant from each version of CAnVAS was selected to show their gaze plot to understand the path that was taken when looking at this task. These two participants were chosen for a representation of gaze plots because they spent almost the same amount of time, approximately 40 seconds, on this task. The gaze plots for the two participants are shown in Figure 4. Analysis for average TVD for all AOIs included in Task 3.3 is shown in Figure 5. This data was not normalized and shows incorrect location AOIs in red, correct location AOIs in green, and neutral AOIs in blue with a one to one trendline added.

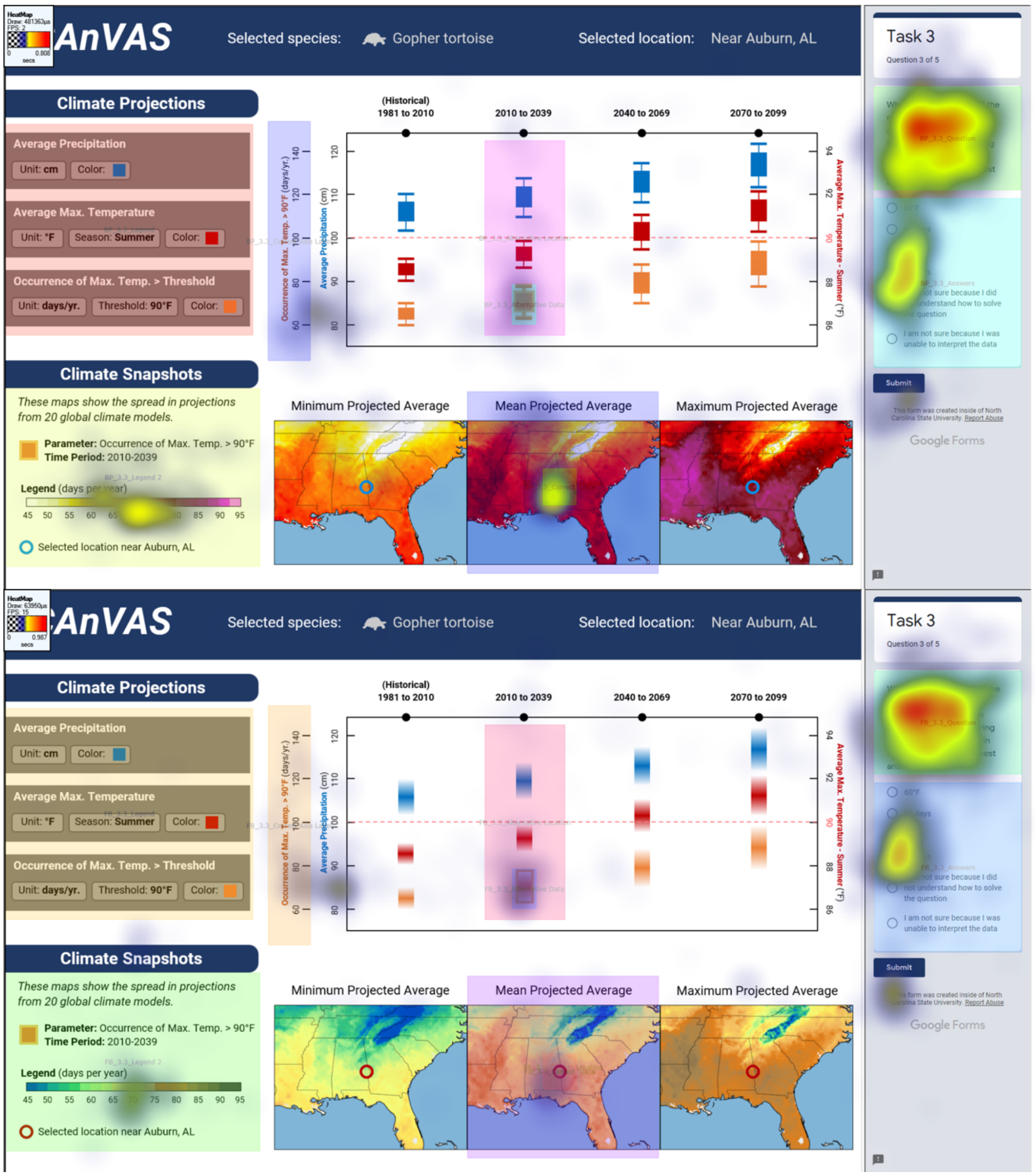

Figure 3: Heatmaps for Task 3.3 with all participants' data.

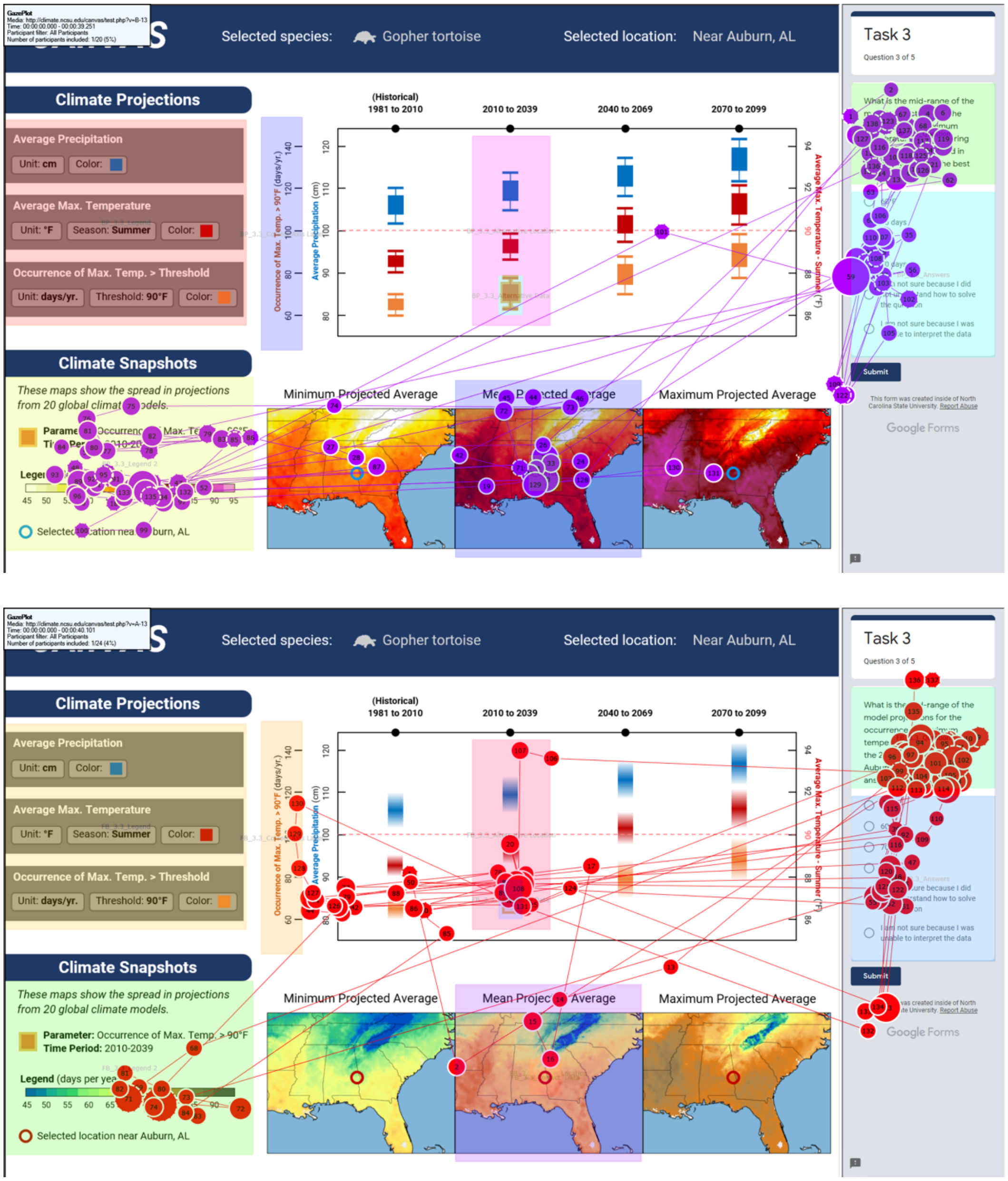

Figure 4: Gaze plots of Task 3.3, for two participants, one from each version of CAnVAS.

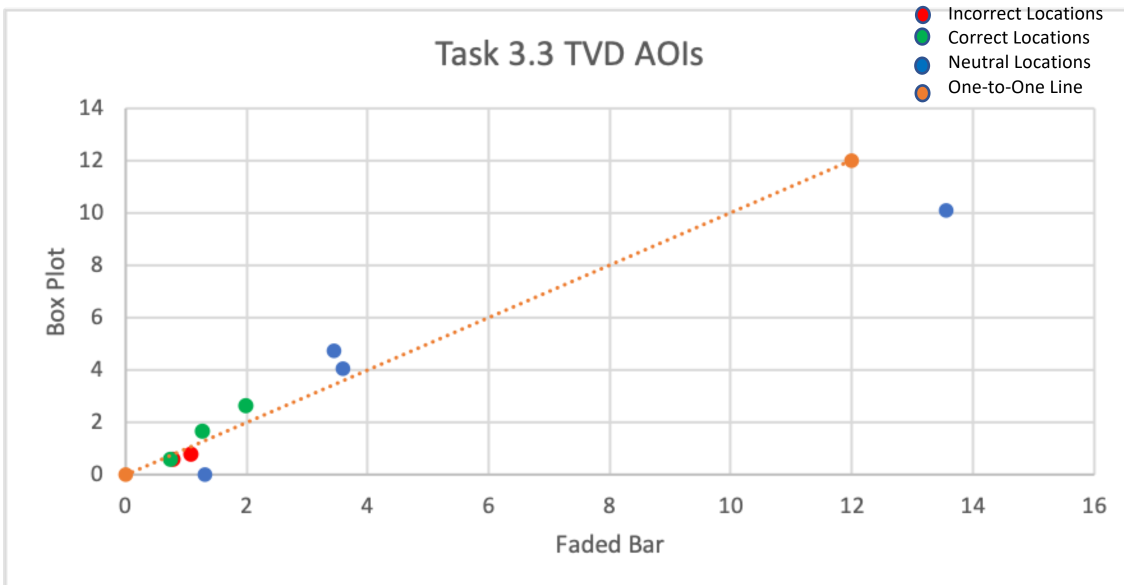

Figure 5: Average Total Visit Duration for all AOIs for Task 3.3

In the heatmaps shown in Figure 3, it is evident that overall, the box plot participants looked at the correct location in the climate snapshot rather than the alternative location in the climate projections. The opposite is true for the faded bar participants as it seems like they used the climate projections rather than the climate snapshots to answer the question. That observation is supported by the same result present in the two gaze plots in Figure 4. The participant in the box plot version strictly looked at the climate snapshots to answer the question. The participant in the faded bar version looked most at the climate projections but did sometimes look at the climate snapshot for the answer. Figure 5 indicates that both the faded bar and box plot participants had no major differences in average TVD for both incorrect and correct AOIs since the data points are close to the one to one trendline.

Eye-Tracking: Commonly Missed Questions

The two most commonly missed questions were Task 1.3 and Task 2.3. In Task 1.3, 12 out of 37 participants answered the question incorrectly and in Task 2.3, 17 out of 37 participants answered the question incorrectly. Both of these questions asked participants to look at a particular piece of data and find the answer on the associated axis label. For this portion of analysis, there are two questions that are addressed by using TFF and TVD metrics: (1) How long did it take for users to look at the correct spot and (2) Did users tend to look at the incorrect post first or correct spot first? TVD is how long the participant visited the particular AOI. This metric does not necessarily mean that the participant has to fixate in the AOI, the participant just would have had to visit the AOI. Correct locations are once again defined as what has been previously established. Incorrect locations would have been an obvious place that participants could have looked to concluded to an incorrect answer. In the case of Task 1.3, the correct location was the average maximum temperature axis label and the incorrect location was the average precipitation axis label. In the case of Task 2.3, the correct location was the occurrence of maximum temperature axis label and the incorrect location was the average precipitation axis label. An independent samples t-test was conducted for each incorrect and correct AOI for both TFF and TVD and it was found that there were no statistical differences between groups. Effect size was measured and TFF Incorrect Location for Task 2.3 was the only AOI with a large effect size (d = 0.52). Tables 5 and 6 below show the descriptive statistics and effect sizes for the commonly missed questions.

Table 5

Analysis of Time to First Fixation for Commonly Missed Questions

Task	**TFF Avg. – FB**	**TFF Std. Dev. – FB**	**TFF Avg. – BP**	**TFF Std. Dev. – BP**	**TFF Effect Size**
1.3 – Correct	0.43	0.33	0.44	0.27	0.03
1.3 – Incorrect	0.43	0.25	0.49	0.20	0.26
2.3 – Correct	0.61	0.46	0.49	0.24	0.35
2.3 – Incorrect	0.54	0.22	0.43	0.23	0.52

Table 6

Analysis of Total Visit Duration for Commonly Missed Questions

Task	TVD Avg. – FB	TVD Std. Dev. – FB	TVD Avg. – BP	TVD Std. Dev. – BP	TVD Effect Size
1.3 – Correct	0.09	0.07	0.10	0.09	0.18
1.3 – Incorrect	0.06	0.05	0.06	0.06	0.03
2.3 – Correct	0.10	0.07	0.10	0.10	0.05
2.3 – Incorrect	0.12	0.09	0.10	0.06	0.30

However, there were qualitative differences which can be seen in heatmaps for these particular questions located in Figures 6 and 7. In these heatmaps, that include all participants for each version, it is evident that some participants looked at the wrong axis labels in order to answer the question. This impacted their answer choice by what axis label the participants looked at. On Figure 6, which displays Task 1.3, there is a concentration of fixations on the average precipitation axis and average maximum temperature axis. On Figure 7, which displays Task 2.3, there is a concentration of fixations on the average precipitation axis and occurrence of maximum temperature axis. Both of these figures show the complications that participants had with these two particular questions due to the multiple y- axes. Figure 8 includes a nonnormalized analysis done for average TVD for all AOIs included on Task 1.3 and 2.3. Figure 8 shows incorrect location AOIs in red, correct location AOIs in green, and neutral AOIs in blue with a one to one trendline added. TVD analysis concluded that there were no major difference in total time visiting both correct and incorrect AOIs for Task 1.3 and 2.3 as the data points were close to the one to one trendline.

Figure 6: Heatmaps for Task 1.3 with all participants' data.

Figure 7: Heatmaps for Task 2.3 with all participants' data.

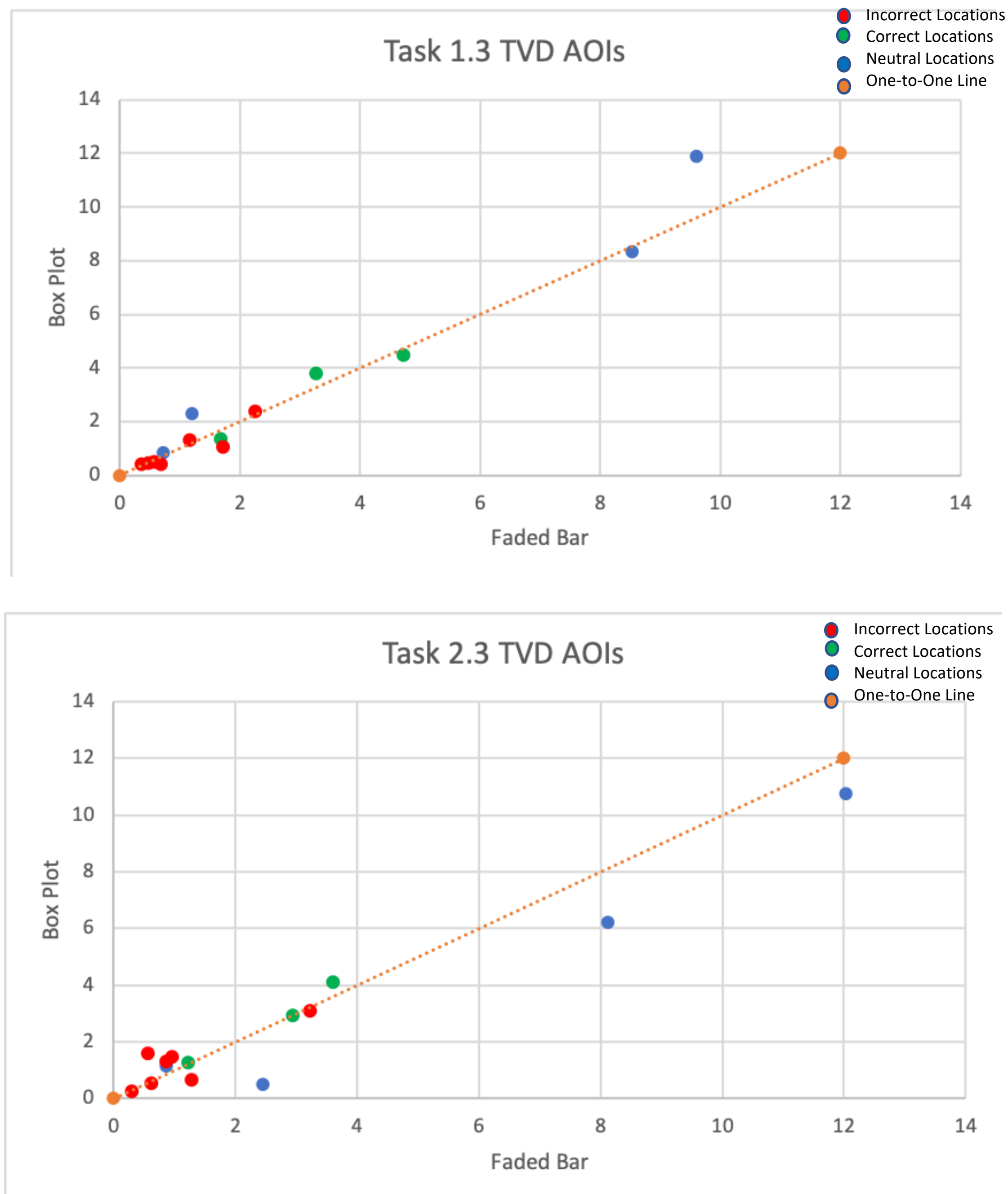

Figure 8: Average Total Visit Duration for all AOIs for Task 1.3 and 2.3

Usability: Efficiency, Effectiveness, Satisfaction

Efficiency was measured using the TFMC eye-tracking metric. TFMC was taken for all 15 questions across both of the treatment groups. TFMC was summed up to test for statistical significance across groups through an independent samples t-test. When looking at the bulk analysis, there was no statistical significance nor large effect size between groups for total TFMC as shown in Table 2. Thus, a finer analysis was done on TFMC through an independent samples t-test with each individual task as shown in Table 7. It should be noted that Tasks 1.1 and 3.1 only registered 13 and 12 participants' TFMC. This more than likely was due to the error in recording for some participants' free explorations which was located right before these two questions. There was also no statistical significance for the TFMC for individual tasks, but averages between groups did vary as can be seen in Table 7. There was only one task, Task 1.5, that had a large effect size for TFMC where faded bar participants took longer to select an answer to the question.

Table 7

Analysis of Time to First Mouse Click on Answer Area of Interest for both versions of Climate Analysis and Visualization for the Assessment of Species Status

Task	**TFMC Avg. – FB**	**TFMC Std. Dev. – FB**	**TFMC Avg. – BP**	**TFMC Std. Dev. – BP**	**TFMC Sig. t-test**	**TFMC Effect Size**
1.1	24.83	12.84	23.92	9.26	0.849	0.08
1.2	27.39	12.68	27.43	9.23	0.990	0.00
1.3	35.11	12.19	36.86	16.40	0.713	0.12
1.4	24.29	13.49	20.65	6.04	0.295	0.35
1.5	33.25	17.11	24.19	14.23	0.090	0.57
2.1	43.34	24.09	44.92	19.06	0.827	0.07
2.2	29.30	19.49	27.25	10.81	0.602	0.17
2.3	34.27	22.32	31.54	17.31	0.681	0.14
2.4	13.53	9.02	9.91	5.49	0.151	0.48
2.5	26.07	14.09	28.56	24.25	0.702	0.13
3.1	36.60	18.66	34.57	10.21	0.741	0.13
3.2	35.22	25.37	31.12	15.24	0.558	0.19
3.3	23.30	21.41	20.87	11.75	0.674	0.14
3.4	64.83	31.69	52.59	36.20	0.281	0.36
3.5	24.29	12.66	25.74	14.42	0.746	0.11

Effectiveness was measured by participants either correctly or incorrectly answering each question. For this study, the best way to represent effectiveness is by showing how many people answered the questions incorrectly. Table 8 shows the results for the effectiveness of the two versions of CAnVAS. Overall, there were 5 questions where both groups had the same number of people answer the questions incorrectly. For the other questions, there were 6 questions where the faded bar participants had a larger number of people answer incorrectly compared to the box plot participants. There were 4 questions where the box plot participants had a larger number of people answer incorrectly compared to the faded bar participants.

Table 8

Number of Participants that Answered Incorrectly

Task	Faded Bar	Box Plot
1.1	1	1
1.2	1	0
1.3	6	6
1.4	2	1
1.5	5	5
2.1	5	4
2.2	0	0
2.3	9	8
2.4	0	0
2.5	2	4
3.1	3	4
3.2	1	2
3.3	0	1
3.4	5	4
3.5	5	4

Note: Shading Indicates Version with Larger Number of Incorrect Answers

Satisfaction was measured by thematic coding of the interviews conducted with undergraduates following their participation in eye-tracking. Table 9 displays the parent codes from the coding process of undergraduate interviews with examples of what those codes would have looked like during interviews. Co-coding with undergraduate interviews was completed for 10% of codes and received good agreement (*Cohen's Kappa* = 0.65).

Table 9

Codes from Analysis of Undergraduate Interviews with Accompanying Excerpts

Code	Excerpt from Interviews
Axis Labels	"…I am not used to looking at an axis label on the righthand side of a graph. I guess that's probably part of the reason. I've never had to interpret a graph that looks this before. Having multiple axes was also challenging, I guess. I think I can have just didn't think about it."
Color	"…Because it was color coordinated, it made it easy for me. Other than that, it wasn't hard to look at. It was easy to find it because it was color coordinated. I could just be like, "All right. This is the color I'm looking at," and then find the numbers beside it."
Correct Interpretation	"Depending on the species, they're each going to have their own needs. They're made to live in certain conditions. If a certain place is becoming unable to inhabit a certain place in a few years, then that's a problem. People are either going to have to try and fix somehow, or they're going to have to relocate in some way."
Difficult to Understand	"At first, [climate projections] was difficult. It took me a second to see which numbers are lining up. Then I even got tricked out on that one question. That was a little bit tricky."
Easy to Understand	"The easiest to immediately get was the precipitation because blue and water. I got confused a couple of times between the average max temperature and the over 90 days. I would say that the snapshots were easier to understand than the red and orange up there."
Incorrect Interpretation	"…I see that the more grayish it is, the hotter it is. Also, the more orangishbrownish, or is the bluer is colder and yellow and green stuff."
Suggestion of Change to CAnVAS	"If I had made the graph, me being me, I probably would have made three separate graphs for each piece of information. One for the precipitation, one for the heat, and one for the occurrence of max temp, but that's just how my brain works. When everything is together, it gets jumbled in my head. If I had separated it out, it would've been easier for me to understand initially."
Task 3	"… At first I was confused at which one I was looking at for the questions. I had assumed it was the bottom one, but for a couple of the first questions I was looking still more at the top chart until I associated it."

With those codes, there were associated child codes indicating satisfaction/dissatisfaction with CAnVAS and even recommendations of changes to CAnVAS. Overall, there were 38 recommendations for changes to CAnVAS. There were 85 times that the code "difficult to understand" was used and only 58 times where the code "easy to understand" was used. The code "axis labels" was frequently used (*n* = 81) and normally this was due to correlating colors of the labels with the data or talking about placements of axis labels.

With the correct and incorrect interpretation codes, there were child codes to distinguish if either version of CAnVAS had a correct or incorrect interpretation. Total, there were 46 times where there was an incorrect interpretation of CAnVAS. On the other hand, there were 79 times where there was a correct interpretation of CAnVAS. Interpretations were either about the version of CAnVAS or describing why species manager would need this tool. Between the two versions of CAnVAS and interpretations, correct and incorrect had a fairly even spread across the two groups. There were slightly more correct interpretations with the box plot (*n* = 27) version of CAnVAS than the faded bar version (*n* = 24). There were slightly more incorrect interpretations of the faded bar version (*n* = 20) versus the box plot version (*n* = 16).

Finally, there were codes specifically about task 3 to understand whether participants actually looked at the climate snapshots or the climate projections of CAnVAS to answer the questions. Task 3 was geared to have participants only look at the climate snapshots to answer the 5 questions associated with that section. Participants identified if they looked only at the climate snapshots (*n* = 7), only at the climate projections (*n* = 11), or if they looked at both (*n* = 26).

DISCUSSION

Species Status Assessments

SSAs, on average, only had less than 3% of climate change information included in their report. Having an explicit section geared specifically towards climate change, for every SSA, will be a step in the right direction of providing accurate information with climate effects on endangered species. Also, the three least used codes in SSAs (e.g., anthropogenic emissions, climate adaptation/resiliency, and uncertainty) show that SSAs are not including as much information on how species are going to adapt to this new environment that is evolving. To be more specific, SSAs do not include as much information about anthropogenic emissions which could influence how people read and understand these documents. By making sure to lay out the consequences of climate change on endangered species, SSAs could be enhanced to better inform managers of such impacts (Corner et al. 2018).

With accuracy, quality and depth, and relevancy, all of these categories could be improved throughout SSAs based on the evidence from ranking of SSAs. Those categories were below the mid-point of the ranking system which acknowledges that advancements need to be made. There were also times where some code categories (i.e. impacts to biology, increase in greenhouse gases, solutions) within the ranking rubric were not included in the text of the assessments. Not including some of those code categories within the SSA does not allow for a robust recognition of the effects that these endangered species could experience. All aspects of species should be considered and acknowledged in SSAs so biologists who may not be familiar with that species will understand exactly what the future of this species looks like under climate change conditions.

Eye-Tracking

Eye-tracking metrics were used to compare the two versions, faded bar and box plot, of CAnVAS. On average, it took the faded bar participants longer to initially fixate on the correct locations implying that participants had a harder time finding where to look for the correct answer. Box plot participants, on average, fixated for longer on most of the correct locations on CAnVAS. This means that for most of the correct locations, the box plot participants might have taken longer to analyze the data being shown. It should be noted that the values for TFD were small in retrospect to the amount of time the participants spent on that particular task. Task 3.3, which had large effect sizes for both TFF and TFD, showed that box plot participants understood to look at the climate snapshots rather than the climate projections to achieve the correct answer. With those two specific eye-tracking metrics, TFF and TFD, the box plot seemed to have results which favor using this version for the final product.

For the commonly missed questions, it was evident, through heat maps, that the axis label location confused participants. Having two axis labels on the left and one on the right impacted how a participant looked at and interpreted the data that was presented. These results are similar to those from Courtney (p. 59, 2019) where the researchers found that there was a challenge with one graph due to requiring participants to look at both the data and the axis label. For Task 1.3, it took box plot participants longer to initially fixate on both the correct and incorrect locations. For Task 2.3, the faded bar participants took longer to initially fixate on the correct location and the incorrect location than the box plot participants did. With TVD, there were no major differences between the two groups, both groups relatively spent the same amount of time visiting the incorrect and correct locations. These commonly missed questions ended up having many participants, across groups, answer incorrectly which is more than likely due to the confusion with the multiple axes rather than preferring one version of CAnVAS over the other. Thus, recommendations revolve around formatting of climate projections rather than choosing a version of CAnVAS.

Usability: Efficiency, Effectiveness, Satisfaction

With efficiency, out of the 15 questions for participants, an average of 10 of those took the faded bar participants longer to select an answer. This could relate to the lack of understanding of the faded bar graphics, thus taking a longer time for participants to realize what they were looking at. Specifically, with the questions in Task 3, 4 of those questions took faded bar participants longer to select an answer than the box plot participants. This might be due to colors of the climate snapshot on the faded bar not necessarily aligning with the data that was being presented of the occurrence of maximum temperature, thus making it harder to understand (Netek et al. 2018). The box plot climate snapshot color scheme also had larger values associated with darker colors which is appropriate for communicating data values (Schloss et al. 2018). From the analysis, the box plot version of CAnVAS was considered to be more efficient than the faded bar version.

The box plot version of CAnVAS proved to be slightly more effective than the faded bar version when looking at correct and incorrect answers. There could be a lot of different factors that pertain to this. For instance, the box plot participants fixated on the correct locations for longer which could have influenced more of the participants answering questions correctly. Also, box plot participants located the correct location faster than faded bar participants which might lead to those participants answering more questions correctly.

With satisfaction, there were more times that participants identified parts of CAnVAS were difficult to understand which could be due to the lack of knowledge in this specific subject. Overall, participants thought the colors of the climate projections on CAnVAS were good for correlating between the legend and the data shown. Most participants were able to correctly interpret which color went with which data set (i.e. red with average maximum temperature, blue with average precipitation, and orange with occurrence of maximum temperature). There were some recommendations during interviews and majority of the recommendations were directed towards the climate projections portion of CAnVAS, suggesting that majority of participants would have liked to see something different with this portion of CAnVAS. Whether it be moving axis labels or breaking up the main graph into 3 separate graphs. Another recommendation by faded bar participants was they indicated that they would have preferred the data be displayed in a different way, perhaps in box plot form. They would interpret the data correctly, but still would want a change that benefitted the usability of the DSS implying they would have preferred box plots. Since there were more correct interpretations than incorrect, the information was largely comprehended by participants.

For Task 3, participants could have found answers on the climate projections for some of the questions, but not all of them. Some participants indicated that they only looked at the climate projections to answer questions which implied that they did not necessarily understand what they should have been looking for within this task. The majority of participants ended up looking at both the climate snapshots and climate projections to answer the questions about task 3. Representations of climate information both spatially (climate snapshots) and temporally (climate projections) were effective at communicating climate information as proven by participants looking at both sections to answer questions which aligns with results in Davis et al.

(2020) even if it was not the end goal of this task to look at both representations. Some participants identified they looked at both since they were already familiar with the climate projections and felt like it aided them to look at the climate projections to understand what exactly was being displayed on the climate snapshots. This result indicates that there could have been a more obvious explanation on CAnVAS of what exactly the climate snapshot was displaying, a visual to direct attention or by having more descriptive text, it could have aided the effectiveness of tasks with undergraduates (Maudlin et al. 2020; Courtney 2019 p. 62). The box plot version of CAnVAS was more satisfying which aligned with higher efficiency and effectiveness with this version (Courtney 2019 p. 58).

This research was able to provide some recommendations on how to improve the CAnVAS DSS. The following includes recommendations to what CAnVAS should look like after this first iteration of the study which will be referred to as CAnVAS 2.0. CAnVAS 2.0 should keep the same colors for displaying the climate projections since participants were able to verbally identify correct interpretations of the colors used. Colors should be related to the data that is displayed and that is evident that the climate projections data does so (Netek et al. 2018). This new version should also have the data displayed as box plots because for most of the tasks, it took faded bar participants longer to fixate on the correct location. Thus, having a box plot display the data may help users better interpret what they are looking at. Also, the box plot participants proved that version of CAnVAS was more efficient and effective compared to the faded bar version. CAnVAS 2.0 should include a more obvious way to show that the data from the climate projections is being pulled down and visualized in the climate snapshots. The color scheme of the climate snapshots should be the coloring of the box plot version. The reasoning for this is it took longer for the faded bar participants to select an answer in 4 questions of Task 3, thus showing a lack in comprehension of what was being shown. The general look of CAnVAS 2.0 should be the box plot version of CAnVAS since it proved to be the best choice of the two versions that were provided. Once again, the box plot version of CAnVAS is shown below in Figure 9 to present a general depiction of CAnVAS 2.0.

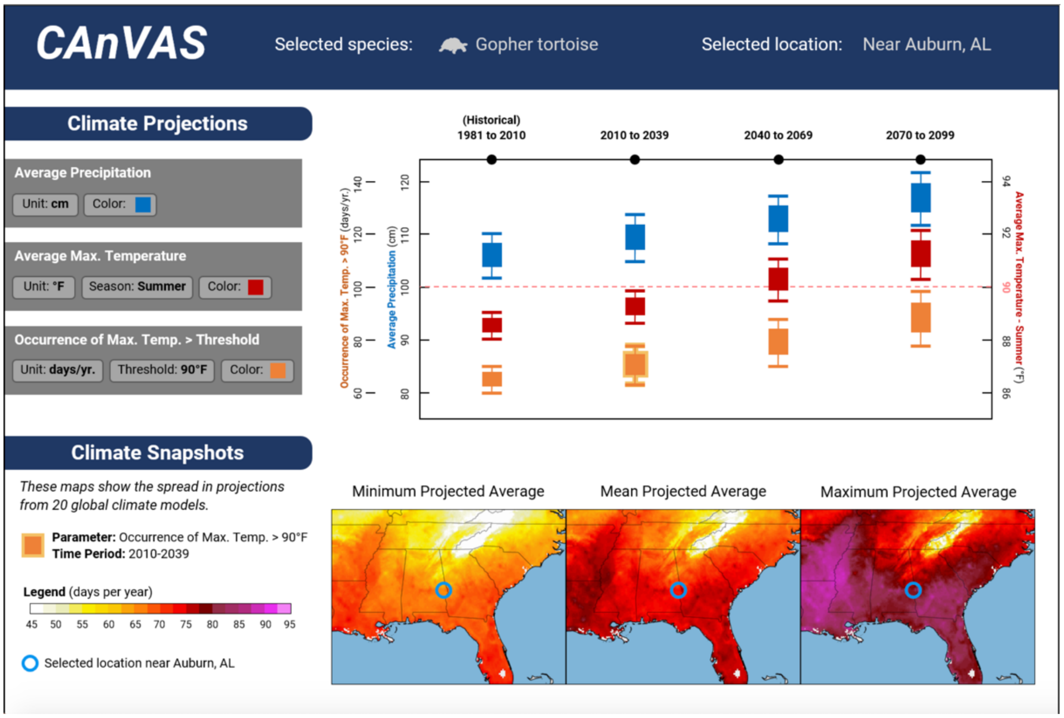

Figure 9: The recommended version of CAnVAS to be used in the future.

CONCLUSION

In this study, co-production has been employed to produce a tool for use by USFWS biologists. Interactions with an initial USFWS committee to assess needs of SSA Authors, led to a first iteration of CAnVAS being developed. Then researchers and developers collaborated to design two versions of CAnVAS to be tested using an A/B study design and a combination of eye-tracking and interview approaches with novice users. Results of this study indicated the box plot version was the best version of CAnVAS. Our recommendations for CANVAS are outlined below:

- Climate projections graphically displayed as box and whisker plots
- Colors of each data set within climate projections should remain the same (i.e. blue for average precipitation, red for average maximum temperature, and orange for occurrence of maximum temperature)
- A redesign of axis labels should be considered
- Climate snapshots use white to purple color scheme for displaying data
- An explanation should be included about how data is being pulled from the climate projections and displayed in climate snapshots

Future Research

Once CAnVAS is refined, instead of being a still image it will be converted to a working website where endangered different species can be selected from a drop-down menu and the climate information for that species will be displayed. This will be beneficial for USFWS biologists since they will be able to look at many endangered species and look at that species' native range or a specified range. After refinement of CAnVAS, more research can be done with this DSS. Future research will consist of eye-tracking stakeholders in the USFWS with the CAnVAS DSS since they are the end user. The results from eye-tracking experts could be compared to the eye-tracking results of novices as well to see difference between groups in future research (Harsh et al. 2019).

After conducting this study with stakeholders, CAnVAS will be able to be refined again in order to benefit the USFWS as much as possible through maintaining collaboration between the science makers and science users. Testing CAnVAS with the end user is important to the DSS development process so that it can be ensured that the USFWS stakeholders correctly interpret the scientific information provided (Davis et al. 2020). An extension of this project is to interview authors of SSAs and analyze those interviews in order to properly assess what climate information they think is needed to properly complete SSAs with accurate climate information. The interviews with authors of SSAs will help to gauge what exactly stakeholders would like to see in CAnVAS. Both of these components will contribute to an improvement in this DSS to ultimately benefit USFWS biologists.

REFERENCES

Alridge, H.D., Davis, C., Boyles, R. (2016). PINEMAP Decision Support System. Acquisition date: 10/29/19. http://climate.ncsu.edu/pinemap/index.php?tool=dev.

Atkins, R.M. & McNeal, K.S. (2018). Exploring differences among student populations during climate graph reading tasks: An eye tracking study. *Journal of Astronomy & Earth Sciences Education*, 5(2), 85-114. https://doi.org/10.19030/jaese.v5i2.10219

Bell, S., Shaw, B., Boaz, A., (2011). Real-world approaches to assessing the impact of environmental research on policy. *Research Evaluation*, 20(3), 227-237. https://doi.org/10.3152/095820211X13118583635792

Blodgett, D., Booth, N., Kunicke, T., Walker, J., Viger, R. (2011). Description and Testing of the Geo Data Portal: A Data Integration Framework and Web Processing Services for Environmental Science Collaboration. USGS Open File Report 2011-1157.

Bojko, A. (2013). Eye Tracking the User Experience: A Practical Guide to Research. New York, NY: Rosenfeld Media.

Climate Explorer. (2016). Available at: http://toolkit.climate.gov/climate-explorer2/

Cho, H., Powell, D., Pichon, A., Kuhns, L. M., Garofalo, R., Schnall, R. (2019). Eye-tracking retrospective think-aloud as a novel approach for a usability evaluation. *International Journal of Medical Informatics*. 129, 366-373. https://doi.org/10.1016/j.ijmedinf.2019.07.010

Corner, A., Shaw. C., Clarke, J. (2018) Principles for effective communication and public engagement on climate change: A Handbook for IPCC authors. Oxford: Climate Outreach.

Courtney, S. (2019) Building Better Graphs for Climate Change Communication: Evidence from Eye-tracking. (Unpublished master's thesis). Auburn University. Auburn, Alabama.

Creager, J. H., & Gillan, D. J. (2016). Toward understanding the findability and discoverability of shading gradients in almost-flat design. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 60, 339-343. https://doi.org/10.1177/1541931213601077

Creswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. Los Angeles: Sage.

Davis, C., Aldridge, H., Boyles, R., McNeal, K. S., Maudlin, L., Atkins, R. (2020). Visually communicating future climate in a web environment. *Weather, Climate, and Society*, 12(4), 877-896. https://doi.org/10.1175/WCAS-D-190152.1

Durose, C., Mangan, C., Needham, C., Rees, J. (2014). Evaluating co-production: pragmatic approaches to building the evidence based. For Co-production Panel, Political Studies Association Conference, April 2014.

Ehmke, C., Wilson, S. (2007). Identifying web usability problems from eyetracking data. Paper presented at the British HCI conference 2007, 03-09-2007 – 07-09-2007, University of Lancaster, UK.

Haesner, M., Chandra, J., Steinert, A., Steinhagen-Thiessan (2018). An eye movement analysis of web usability: Differences between older adults with and without mid cognitive impairment. *Assistive Technology*, 30(2), 66-73. https://doi.org/10.1080/10400435.2016.1251994

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W., Medina-Elizade, M. (2006). Global temperature change. *Proceeding of the National Academy of Sciences*, 103, 14288-14293. https://doi.org/10.1073/pnas.0606291103

Harold, J., Lorenzoni I., Shipley, T. F., Coventry, K. R. (2016). Cognitive and psychological science insights to improve climate change visualization. *Nature Climate Change*, 6, 1080-1089. https://doi.org/10.1030/nclimate3162

Harsh, J. A., Campillo, M., Murray, C., Myers, C., Nguyen, J., Maltese, A. V. (2019). "Seeing" Data Like an Expert: An Eye-Tracking Study Using Graphical Data Representations. *CBE – Life Sciences Education*, 18 (3). https://doi.org/10.1187/cbe.18-06-0102

Hegarty, M., Canham, M. S., Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 36(1), 37-53. https://doi.org/10.1037/a0017683

Hulme, P. E., (2005). Adapting to climate change: Is there scope of ecological management in the face of a global threat? *Journal of Applied Ecology*, 42, 784-794. https://doi.org/10.1111/j.1365-2664.2005.01082.x

Jacob, R. J., Karn, K. S. (2003). Eye-tracking in human-computer interaction and usability research: ready to deliver the promises. In *The mind's eye,* 573-605. North-Holland.

King, A. J., Bol, N., Cummins, R. G., John, K. K. (2019). Improving visual behavior research in communication science: An overview, review, and reporting recommendations for using eye-tracking methods. *Communication Methods and Measures*, 13(3), 149-177. https://doi.org/10.1080/19312458.2018.1558194

Maudlin, L. C., McNeal, K. S., Dinon-Aldridge, H., Davis, C., Boyles, R., Atkins, R. M. (2020). Website usability difference between males and females: An eye-tracking evaluation of a climate decision support system. *Weather, Climate, & Society*, 12(1), 183-192. https://doi.org/10.1175/WCAS-18-0127.1.

Mason, S. M., Kne, L., Dyke, K. R., Shannon J., Eria, S. (2012). Using eye-tracking and mouse metrics to test usability of web mapping navigation. *Cartography and Geographic Information Science*, 39(1), 48-60. https://doi.org/10.1559/1523040639148

McNie, E. C. (2012). Delivering climate services: Organizational strategies and approaches for producing useful climate-science information. *Weather, Climate, and Society*, 5(1), 14-26. https://doi.org/10.1175/WCAS-D-11-00034.1

Meadows, A. M., Ferguson, D. B., Guido, Z., Horangic, A., Owen, G., Wall, T. (2015). Moving toward the deliberate coproduction of climate science knowledge. *Weather, Climate, & Society,* 7(2), 179-191. https://doi.org/10.1175/WCAS-D-14-00050.1

Moreland, K. (2009). Diverging color maps for scientific visualizations. In: Bebis G. et al. (eds) Advances in Visual Computing. ISVC 2009. Lecture Notes in Computer Science vol. 5876. Springer, Berlin, Heifelber. https://doi.org/10.1007/978-3-642-10520-3_9

Netek, R., Pour, T., Slezakova, R. (2018). Implementation of heat maps in geographical information system – exploratory study on traffic accident data. *Open Geosciences*, 10(1), 367-384. https://doi.org/10.1515/geo-2018-2019

Nisbet M. C. (2009). Communicating climate change: Why frames matter to public engagement. *Environment*, 51, 514-518. https://doi.org/10.3200/ENVT.51.2.12-23

Okan, Y., Galesic, M., Garcia-Retamero, R. (2016). How people with low and high graph literacy process health graphs: Evidence from eye-tracking. *Behavioral Decision Making*, 29 (2-3), 271-294. https://doi.org/10.1002/bdm.1891

Pyke, C.R., Bierwagen, B.G., Furlow, J., Gamble, J., Johnson, T., Julius, S., West, J., (2007). A decision inventory approach for improving decision support for climate change impact assessment and adaptation. *Environmental Science & Policy*, 10, 610-621. https://doi.org/10.1016/j.envsci.2007.05.001

Schloss, K. B., Gramazio, C. C., Silverman, A. T., Parker, M. L., Wang, A. S. (2018). Mapping color to meaning in colormap data visualizations. *IEEE transactions on visualization and computer graphics,* 25(1), 810-819. https://doi.org/10.1109/TVCG.2018.2865147

Schwandt, T. A. (2015). The SAGE dictionary of qualitative inquiry (4th). Thousand Oaks, CA: Sage.

Shah, P., Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, 3 (3), 560-578. https://doi.org/10.1111/j.1756-8765.2009.01066.x

Somerville, R. C. J., Hassol, S. J. (2011). Communicating the science of climate change. *Physics Today*, 64(10), 48-53. https://doi.org/10.1063/PT.3.1296

Turton, T. L., Ware, C., Samsel, F., Rogers, D. H. (2017, June). A crowdsourced approach to colormap assessment. In *EuroVis Workshop on Reproducability, Verification, and Validation in Visualization (EuroRV#).* The Eurographics Association.

USGCRP, (2018). *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II* [Reidmiller, D. R., C. W. Avery, D. R. Easterling, K.E. Kunkel, K.L.M. Lewis, T.K. Maycock, and B.C. Stewart (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, 1515 pp. https://doi.org/10.7930/NCA4.2018

VanderMolen, K., Meadow, A. M., Horangic, A., Wall, T. U. (2020). Typologizing stakeholder information use to better understand the impacts of collaborative climate science. *Environmental Management*, https://doi.org/10.1007/s00267-019-01237-9

Wall, T. U., Meadow, A. M., Horganic, A. (2017). Developing evaluation indicators to improve the process of coproducing usable climate science. *Weather, Climate, and Society*, 9(1), 95-107. https://journals.ametsoc.org/view/journals/wcas/9/1/wcas-d-16-008.xml

APPENDIX A

Coding Scheme Used for SSAs:

Codes will be identified as clauses instead of paragraphs. This will remain consistent throughout the duration of coding. Below I have provided some definitions of what these codes will look like in the SSAs. The following is the thematic code with parent and child codes:

- *Climate Change*: Any mention of climate change
- *Extreme Weather:* This is an overall general code for if any type of extreme weather is mentioned. The child codes below will be used if they are explicitly mentioned in the SSA.
    - *Drought:* Drought is specifically mentioned in some SSAs. In some instances, this may also be coded with Increase in Temperature (see below).
    - *Hurricanes/Tropical Storms/Tropical Depressions:* Any specific mention of hurricanes, tropical storms, or tropical depressions.
    - *Wildfires:* Any specific mention of wildfires. In some instances this may be co-coded with Increase in Temperature or Precipitation Pattern Changes (see below).
- *Impacts to Biology:* This is an overall general code for if any type of impacts to biology is mentioned. The child codes below will be used if they are explicitly mentioned in the SSA.
    - *Deforestation:* Any mention of forest reduction impacts.
    - *Sea Level Rise:* Any mention of sea level impacts.
    - *Urbanization:* Mention of urbanization impacting a certain species. This potentially could be co-coded with deforestation.
    - *Storm Surge:* Any mention of storm surge impacts.
    - *Tidal Flooding:* Any mention of tidal flooding impacts.
    - *Habitat Fragmentation:* Any mention of habitat fragmentation. This potentially could be co-coded with urbanization.
- *Inaccuracies:* Any noted inaccuracies with any of the listed codes.
- *Increase in GHGs:* This is an overall general code for if increase in GHGs is mentioned. The child code below will be used if they are explicitly mentioned in the SSA.
    - *Anthropogenic Emissions*: Mention of anthropogenic causes of GHGs (i.e. burning of fossil fuels, cement manufacturing)
- *Increase in Temperatures:* Any mention of increased temperatures.
- *Precipitation Pattern Changes:* This code will be used when the SSA explicitly mentions precipitation pattern changes or mentions the impact that temperatures and droughts have on precipitation. This could be co-coded with Increase in Temperatures or Drought.
- *Solutions:* This is an overall general code for if any type of solutions are mentioned. The child codes below will be used if they are explicitly mentioned in the SSA.
    - *Climate Adaptation/Resiliency*: What will be done to help the species react to the changing climate effects or how might the species be able to adapt naturally.
- *Use of Graphics*: The direct presence of climate related graphics in the SSAs
    - *Future Projections*: Graphics that show future projections involving climate
    - *Uncertainty*: Any graphics that mention uncertainty

APPENDIX B

Ranking Rubric Used for SSAs:

Accuracy

1: The information provided on the given topic was completely inaccurate pertaining to the impact this code will have on species.
3: The information provided was somewhat accurate, but contained inaccuracies pertaining to the impact this code will have on species.
5: The information provided was completely accurate pertaining to the impact this code will have on species.

Quality & Depth

1: There was inadequate information present to provide a connection between that code and the impact it will have on the specific species.
3: Information was present but did not fully describe the impact that code will have on the specific species.
5: Adequate information was provided to fully describe the impact it will have on the specific species.

Relevancy

1: The information provided did not pertain to how the species will be affected by this code. Thus, did not contribute to the overall learning of the impacts of this code.
3: The information provided was somewhat pertinent with how the species will be affected by this code.
5: The information provided was completely pertinent with how the species will be affected by this code.

N/A: Code was not mentioned in the SSA.

	Accuracy	**Quality & Depth**	**Relevancy**
Climate Change			
Extreme Weather			
Impacts to Biology			
Increase in GHGs			
Increase in Temperature			
Precipitation Pattern Changes			
Solutions			
Use of Graphics			

APPENDIX C

Eye-Tracking Interview Questions:

1. Explain how you navigated the tool and provide your interpretation (when appropriate) of each of the following characteristics of the tool below?
    a. How were you able to identify the highest possible value and lowest possible (e.g., T or precipitation, days >90) value after plotting the different climate projections? Explain what "multi-model mean" means to you.
    b. How were you able to identify which axis label went with each climate projection? What are the appropriate units for each of the variables shown?
    c. How did you interpret the different colors that were used for plotting the climate projections?
    d. How were you able to get the tool to map the location you wanted to explore? Did you prefer either the native range information or the specific location information? Why/Why not?
    e. How did you interpret the [faded colored bars] or [bars at the top and the bottom] of the shown data trends over time to mean?
2. Why do you think T, P, and days over 90 degrees are important information for species managers to understand?
3. What components of this tool were difficult to understand and/or navigate? Why?
4. What components of this tool were easy to understand and/or navigate? Why?
5. If you could change anything about this tool, what would you change and why?

APPENDIX D

Coding Scheme for Interviews with Eye-Tracking Participants:

- ***Correct Interpretation:*** Any correct interpretation that does not pertain to the child codes below
    - ***Box and Whisker:*** Participant was able to accurately describe the box and whisker plot
    - ***Faded Bar:*** Participant was able to accurately describe the faded bar plot
    - ***Why This is Important for Species Managers:*** Participant was able to correctly describe how climate change will impact species
- ***Incorrect Interpretation:*** Any incorrect interpretation that does not pertain to the child codes below
    - ***Box and Whisker:*** Participant was not able to accurately describe the box and whisker plot
    - ***Faded Bar:*** Participant was not able to accurately describe the faded bar plot
    - ***Why This is Important for Species Managers:*** Participant incorrectly described how climate change will impact species
- ***Suggestion of Change to CAnVAS***
    - ***Change to Climate Projections:*** Participant mentions wanting to change something involving the climate projections (may say something like "top portion" or "top graph")
    - ***Change to Climate Snapshots:*** Participant mentions wanting to change something involving the climate snapshots (may say something like "bottom portion" or "bottom graph")
    - ***General CAnVAS Change:*** Participant mentions a general change about the CAnVAS website not pertaining to the Climate Projections or Climate Snapshots
- ***Difficult to Understand:*** Anything that the participant felt like was difficult to understand with the tool; also participant not recognizing something is there
- ***Easy to Understand:*** Anything that the participant felt like was easy to understand with the tool
- ***Color***: Any general mention of color
    - ***Linking Climate Projections:*** Participant interpreted the different colors that were used for the climate projections (may be described as "matching")
    - ***Satisfaction with Climate Snapshot:*** Participant was satisfied with colors used for Climate Snapshot (may be "thought it was easy" or "liked the colors")
    - ***Dissatisfaction with Climate Snapshot:*** Participant was dissatisfied with colors used for Climate Snapshot (may be "thought it was difficult" or "did not like the colors")
- ***Axis Labels:*** Any mention of axis labels from the participants/may be used only as a categorial code
    - ***Placement of Labels:*** Participant describes confusion with the placement of axis labels (may be "having two on one side is hard" or "used to looking to the left")
    - ***Intervals of Labels:*** Participant describes confusion with the different intervals with each axis label

- ***Task 3***
  - ***Strictly Looked at Bottom Graph:*** Any indication that the participant looked only at the Climate Snapshot in order to answer any questions on Task 3
  - ***Strictly Looked at Top Graph:*** Any indication that the participant looked only at the Climate Projections in order to answer any questions on Task 3
  - ***Looked at Bottom and Top Graphs:*** Any indication that the participant looked at both the Climate Snapshots and Climate Projections in order to answer any questions on Task 3

APPENDIX E

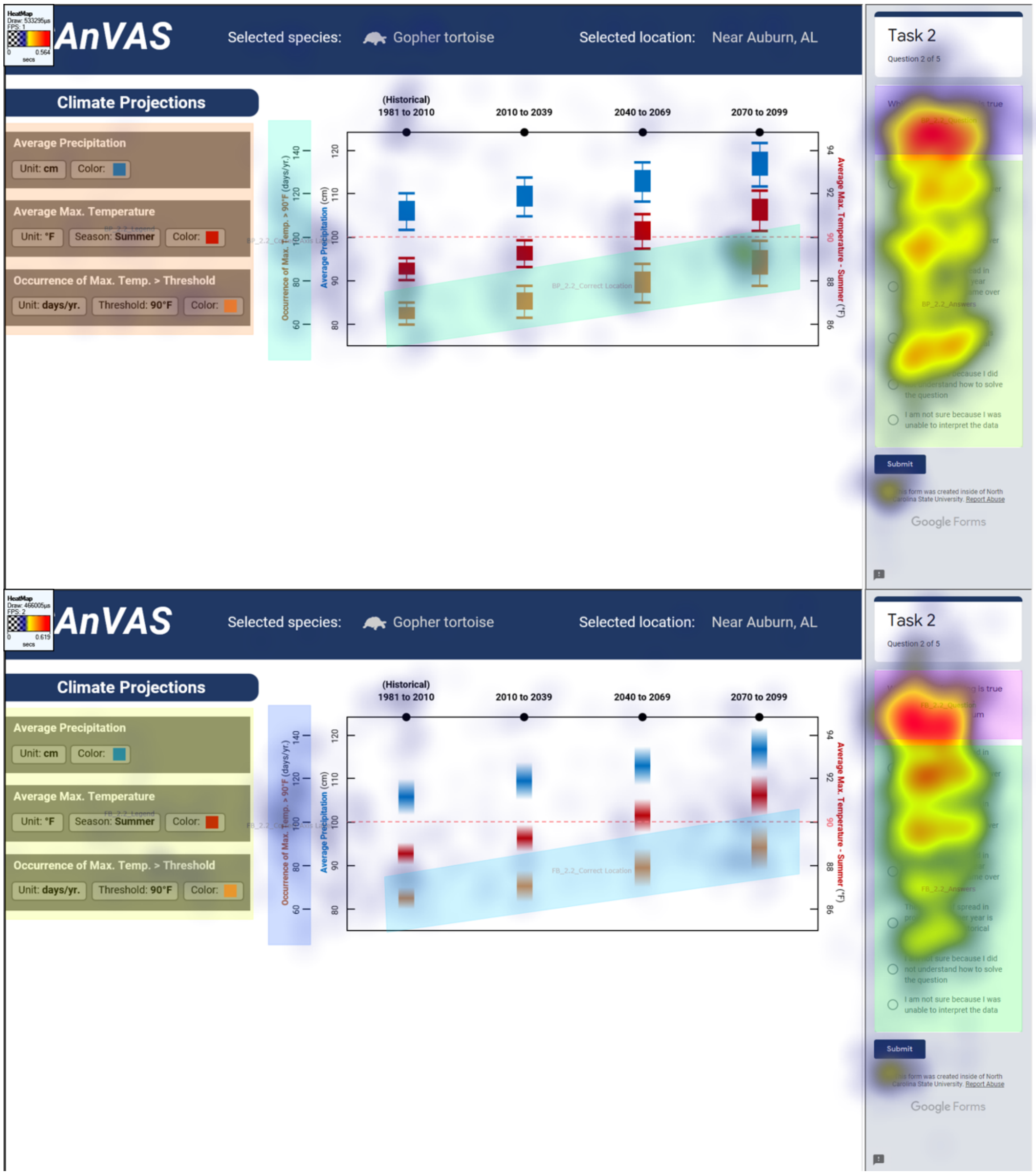

Figure 10: Heatmaps for Task 2.2 with all participants' data.

Figure 11: Heatmaps for Task 2.4 with all participants' data.

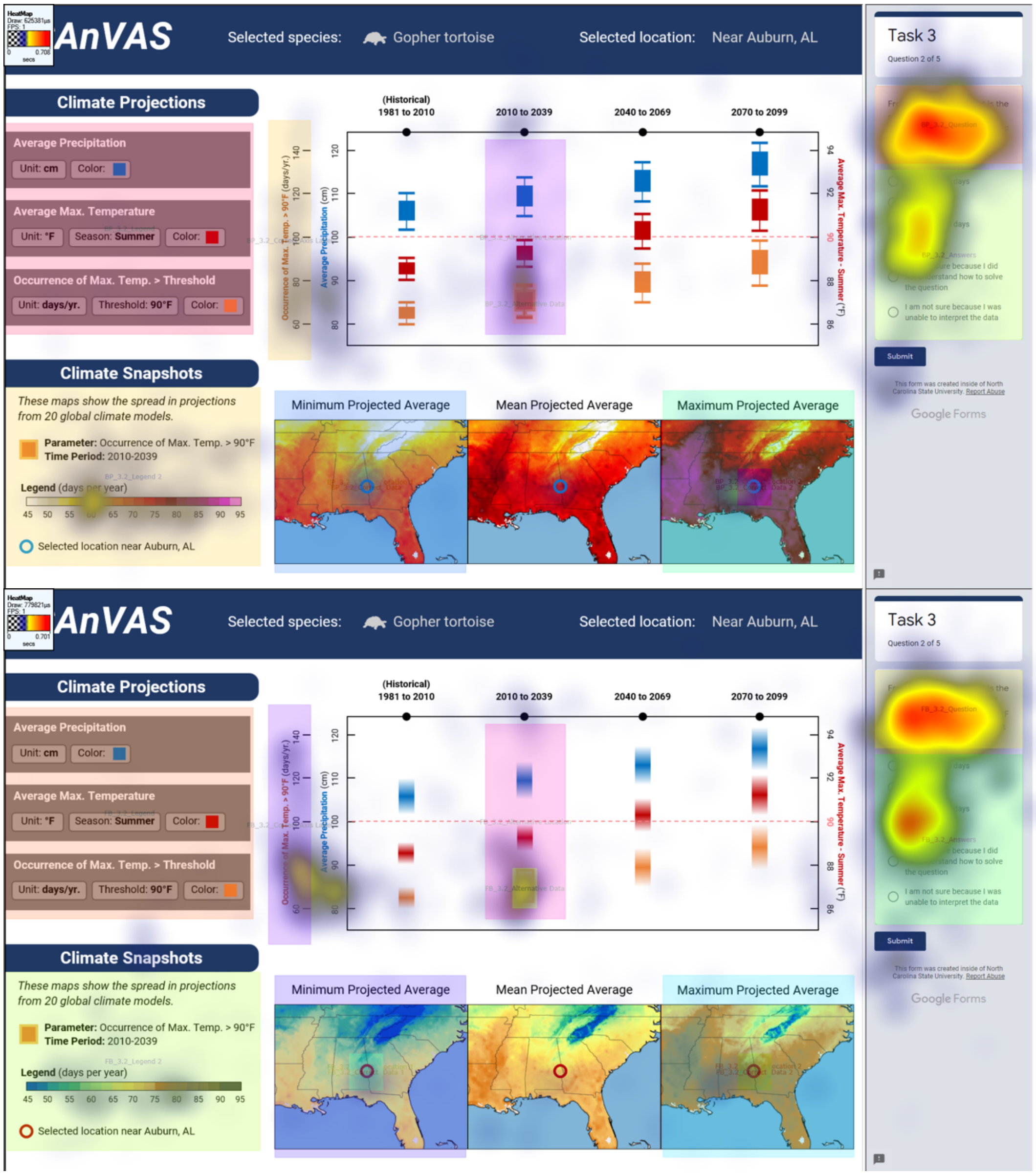

Figure 12: Heatmaps for Task 3.2 with all participants' data.

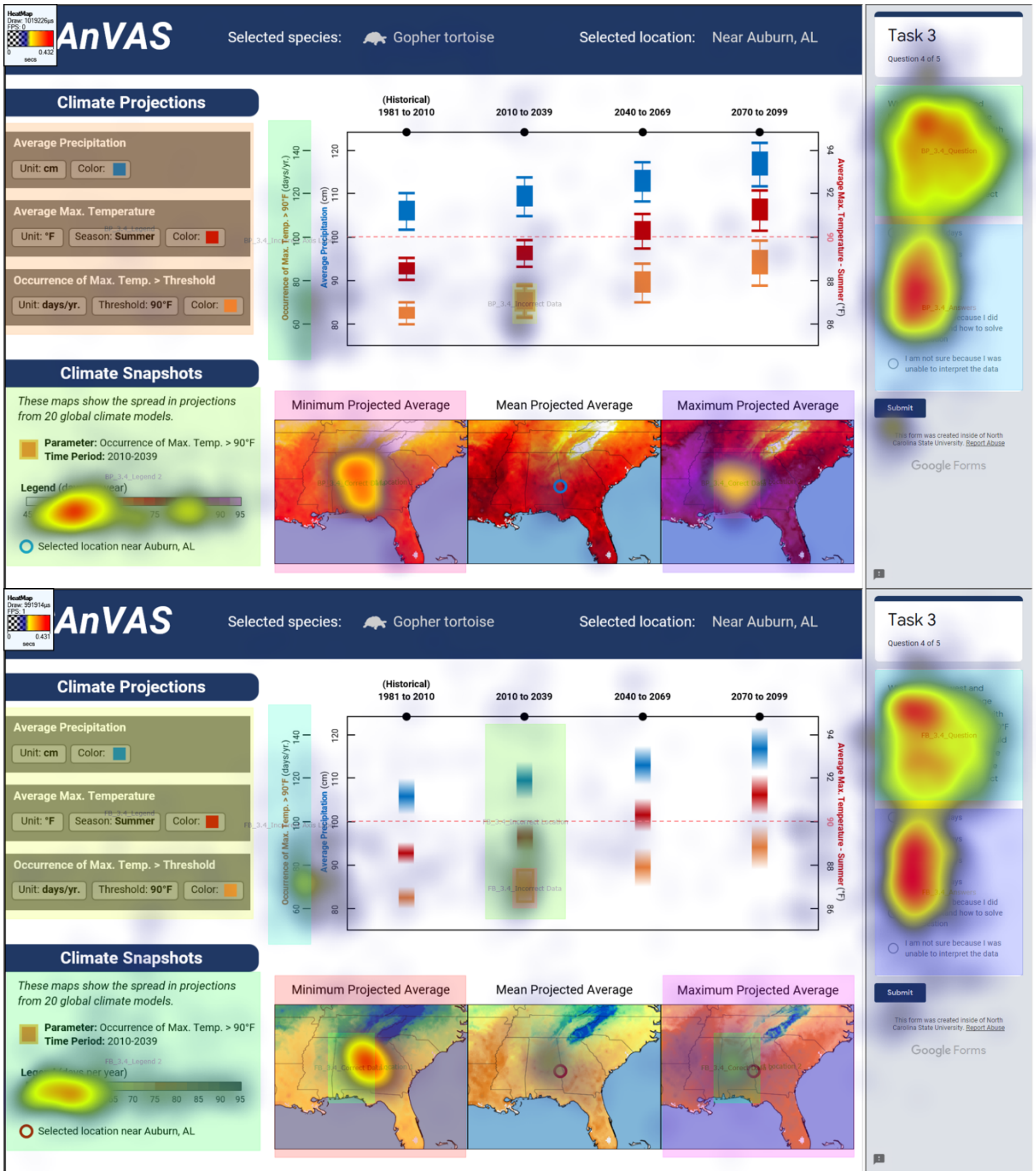

Figure 13: Heatmaps for Task 3.4 with all participants' data.

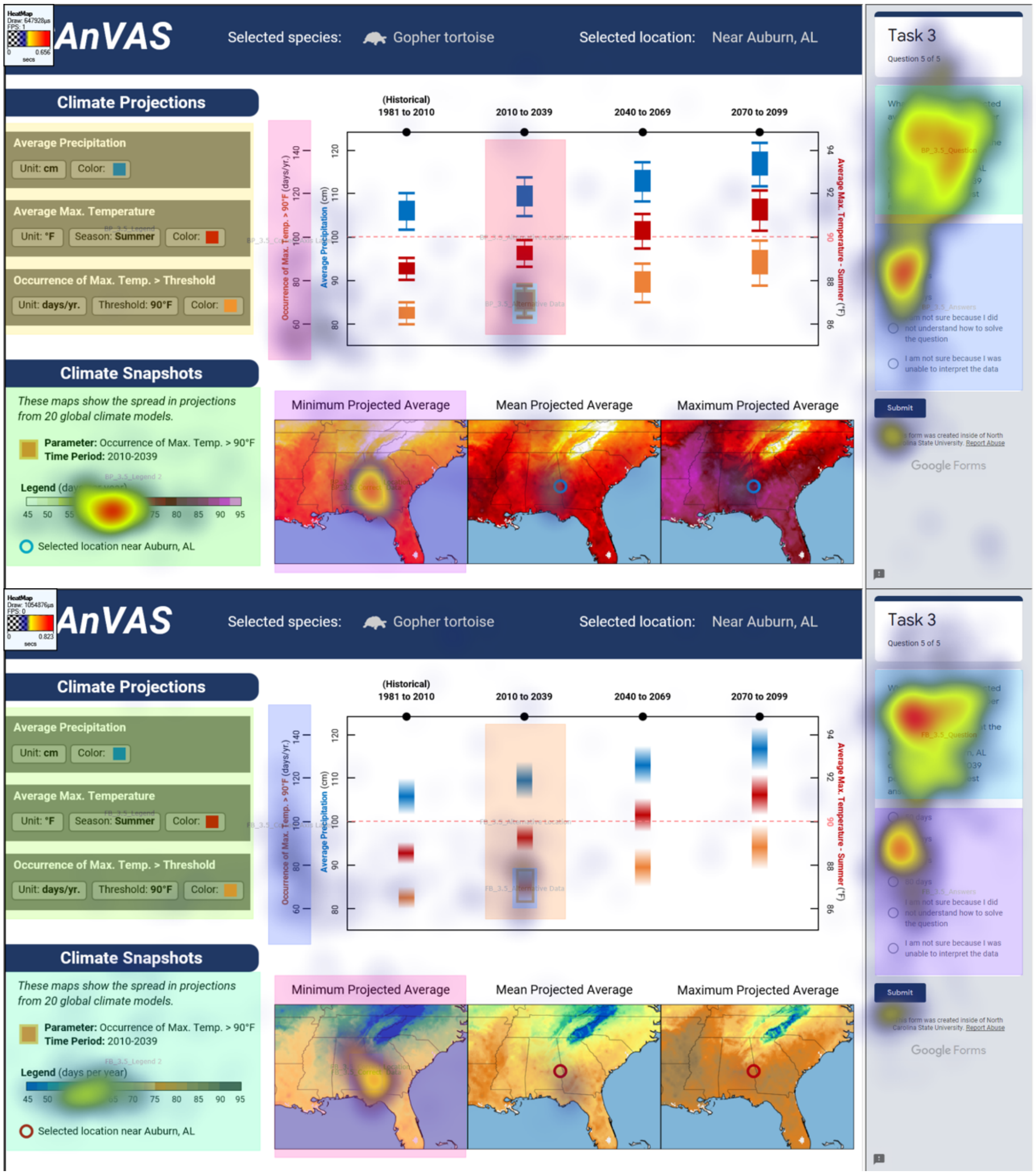

Figure 14: Heatmaps for Task 3.5 with all participants' data.