Tracking Eye Movement When Observing Statistical Graphics

by

Amanda Murdoch

A research paper
presented to the University of Waterloo
in partial fulfillment of the
requirement for the degree of
Master of Mathematics
in
Computational Mathematics

Supervisor: Prof. R. Wayne Oldford

Waterloo, Ontario, Canada, 2014

I hereby declare that I am the sole author of this report. This is a true copy of the report, including any required final revisions, as accepted by my examiners.

I understand that my report may be made electronically available to the public.

Abstract

This project is an investigation of where people look when they make simple decisions about statistical graphics. The goal is to learn how different people's eye path may be when they observe the same image, and answer a specific question. A Tobii eyetracker was used, and a slideshow application integrated with eye tracking software to record the path of people's eyes on the screen during each slide. Plots were created to test a variety of features in plots, and plots used included scatterplots, boxplots and Chernoff faces. Participants were recruited, and data gathered on their eye path, as well as answers to the questions asked while they viewed the images, and some demographic information. A subset of the data is then presented here, showing the eye path as an overlay on the image shown. Based on the images it is shown that some people seem to take a more methodical approach than others, and people's paths seem very different depending on the plot shown. More analysis is needed to determine any relationship between the answers to the questions asked and the eye track, and there is lots of potential for further research in this area.

Acknowledgements

I would like to thank my supervisor, Professor R. Wayne Oldford, for being just as enthusiastic about this project as I was, and providing great advice for all my questions. Thanks also goes to everyone who ever sat in front of the tracker, before and during the study, and gave me great feedback about improvements to make, Python help, and finally data to work with.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The goal of this project is to investigate where people look when viewing statistical graphics. There is a big question of whether different people looking at the same image with the same question in mind will observe the same key elements, and come to the same conclusion. By using an eye tracking device when presenting a slideshow of images, each with a simple associated question, we have measured the path of the person's eyes on the screen. Three types of plots were displayed for participants, each with a different common question. The bulk of this project involved setting up the eye tracking device and slideshow software, and generating a variety of appropriate plots to use in the slideshow. Once everything was set up, and ethics clearance procured, 21 participants were recruited to participate in the experiment. Some results from those samples are presented here, leading to some preliminary conclusions about how people view and make decisions about plots.

1.1 Background Information and Prior Research

The eye provides us with vision through a combination of structural properties. The cornea, the outside layer of the eye, is shaped so that light refracts towards the back of the eye. The pupil, the black center, changes in size in order to allow the right amount of light to refract. The iris, the coloured part surrounding the pupil, controls the changes in the pupil size. Once light passes through the front of the eye, it is registered using the retina at the back of the eye, where the information is then passed to the brain for interpretation. As a result of the complete anatomy of the eye, we have detailed central vision and blurry peripheral vision. The less detailed peripheral vision often brings our attention to objects worth more investigation that lay outside of the central vision, leading to examination directly. Using an eye tracking system, the available data describes where the pupil is pointed and hence approximately where the central vision is directed, but can not tell us about what a person notices peripherally. The point where the eyes are directed and focused on (so in this study the point on the computer screen the eye is directed at) is called the gaze point. The hope given a study such as this is that people will always investigate relevant details directly, but they may also use an unknown amount of information gained using peripheral vision. The human brain uses many strategies of classifying objects and patterns based on a quick glance, so it is not expected that a person would direct their eyes to every detail that they register.

Previous studies using eye tracking have been used for a variety of applications, including for understanding the human brain, testing web design, marketing, and many more. One early study tests how the question asked about an image effects the direction of the gaze point on a plot (Yarbus, 1967). See Figure 1.1 for details.

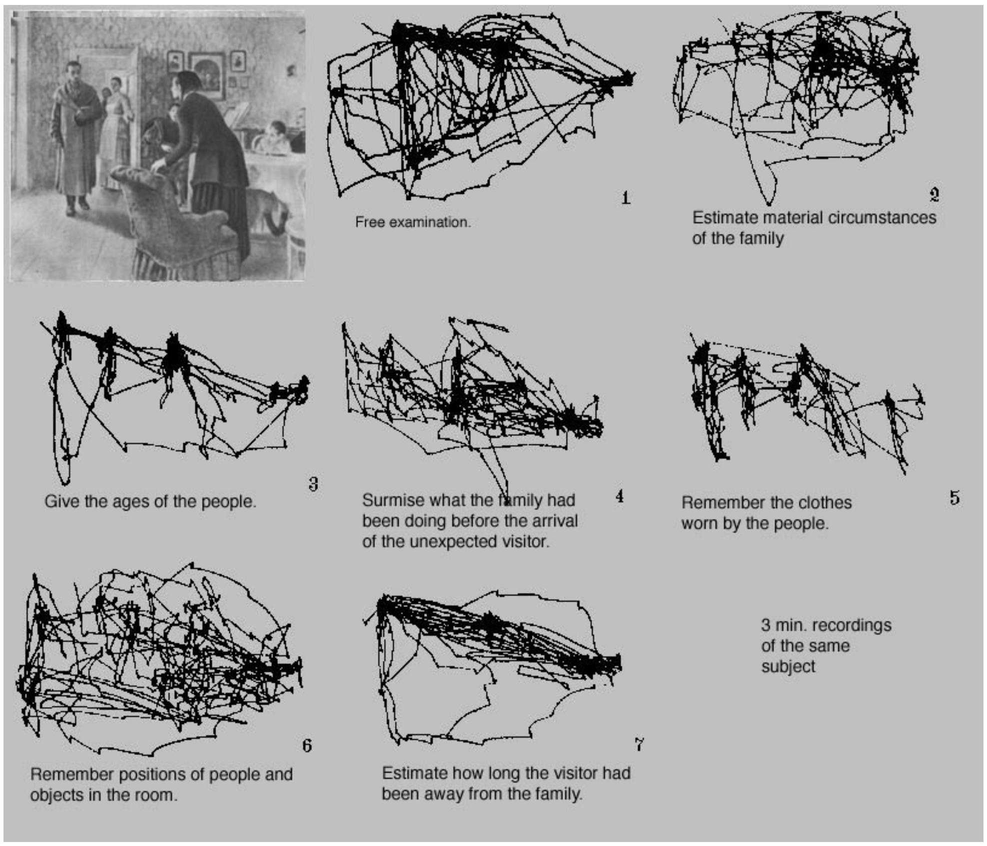

Figure 1.1: Demonstration of the question that is asked changing the type of gaze path seen. The first image (top left) was shown to participants, and the other images show one participant's gaze path when asked each of the different questions. (Yarbus, 1967)

Chapter 2

Eye Tracking

2.1 Eye Tracking with Tobii

A Tobii X2-30 Compact Eye Tracker has been used to track eye movements over time. The tracker is a small device that attaches to the bottom of a monitor or laptop screen, and given a person is within range, both eyes are detected and data is recorded. Data received from the tracker includes the eye's position in space, the location on the screen the eye is directed at, pupil diameter and more. The frequency of measures is about 30Hz, though sometimes interruptions lead to less data. The measurement used to detect the eye is a Pupil Centre Corneal Reflection technique (PCCR), where essentially a light is shone from the device and eyes are detected based on pattern recognition in the reflection of the light. Both the pupil and cornea locations are known based on the reflection, which leads to a vector describing the direction of gaze. In order to direct data to a user interface, a few sample programs were included with the eye tracker which were then adapted to the needs of this project. All coding for this project was done using Python.

Since every user has individual characteristics, a calibration procedure is used before tracking. Calibration creates a model of the individual's eye based on the characteristics of the reflections recorded. When calibration is started the user is instructed to look at dots as they appear on the screen. The location of the dots is set the same for all users, so that the model can match the recorded pupil and cornea location to where they are looking on screen. An image showing the quality of the calibration is shown after, with vectors for each calibration dot showing the difference between the user's gaze point and the calibration dot.

2.2 Creation of a Plot Slideshow Eye Tracking Program

Of the sample programs included, the most similar program to our needs is an application that calibrates the eye tracker to the user and then shows the approximate relative location of the eyes in space. Using that program as a starting point, the major items to change were to record the gaze point on the screen rather than displaying the eye location in space, and to create a slideshow option with gaze recorded individually for each slide. Since the eye tracking software makes the gaze point on the screen as easily available as the eye's location, switching from one the other was quite simple. Gaze point is recorded as numeric values, to 12 decimal digits. After calibration, the value of the gaze point corresponds to the location of the gaze on the monitor, scaled so both horizontally and vertically the range of gaze points within the screen is $[0, 1]$ Accessing the gaze point to output to file was a matter of extracting the eye tracker's information and then changing the variable type to be a string to output to file.

The slideshow is integrated with the eye tracking code so that when the slideshow begins a new window is opened, a timer runs, and the image is changed when the timer runs out. The window that opens has a widget in it containing the image to be shown, and the widget is reprinted every time the image changes. At first a Python 'image' widget was used, but that interfered with the eye tracking so that no information was printed to file, so a Python 'pixbuf' drawable image widget was then used, solving the interference problem. Since during calibration a new window is opened while the tracker is recording data, that provided code to start with for the slideshow. During the slideshow, output files are saved in folders according to the section of the study and participant number, for ease of use later.

2.3 Data Challenges

Some challenges are associated with the eye tracking data, that lead to questions of how to best represent a gaze path. When gaze point is recorded, the left and right eye each have a point. Given an accurate calibration and normal conditions, the left and right eye show differences in gaze point. For users that seem to not calibrate as well, or with changing environmental conditions (the user's location, amount of moisture in the eye, *etc.*), the difference between the two eyes grows. There are also some times where the two eyes are measured to be looking in very different places for just a moment in time. These moments seem to be errors in measure, but there is no way to tell what may have happened at that time to lead to a discrepancy, and whether that result is worth recording. It would be possible to average the two eye's gaze locations to get an approximate location where the person is trying to look, but since that may remove important information from the data, both tracks are shown for this study.

Another challenge with the data is that the gaze is measured for some distance outside of the monitor, which leads to values outside of the $[0, 1]$ boundary of the coordinates of the edges of the screen. The measures outside the screen should be still accurate as long as they are recorded, but it does pose a problem for presentation, since the values outside what appears on the screen may be cut off.

Chapter 3

Generating Plots

Three types of plots were used in this study, scatterplots, matrices of plots, and chernoff faces. Of the scatterplots most were generated using the algorithms shown below, and two were previously used having been generated using distance data. The Chernoff faces were entirely generated using the R iris dataset. The matrices of plots were also used previously for testing a slightly different problem, in identifying how good people are at choosing the real data among images randomly generated to look similar.

3.1 Generating scatterplots with structure

Given the goal of finding which features of graphs people look at the most, data was generated both with no underlying structure and with some known underlying structure. Structure of interest includes clumps of points, points forming lines within the plot, and extra whitespace being added. A scatterplot with no underlying structure can be created by sampling from a uniform distribution in two variables. A resampling technique was used to create clumpy data or lines in the data, using the algorithm as follows:

Algorithm 1:

1. Sample $(x_1, y_1), ..., (x_n, y_n)$ from Unif(0, 1),
2. For each sampled point i, $i \in 1, ..., n$, select an error parameter in each direction (σ_{xi} and σ_{yi}) and a rotation parameter (θ_i)

3. Resample m points from the original points, with replacement, maintaining the associated error and rotation parameters

4. Generate a jittering value for each resampled point using the normal distribution and error parameters, such that $\delta_{xj} \sim \text{Normal}(0, \sigma_{xj})$ and $\delta_{yj} \sim \text{Normal}(0, \sigma_{yj})$ for resampled point j, where $j \in 1, ..., m$,

5. Rotate each pair of jittering values using a rotation matrix.

$$\begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix}$$

In order to create plots with extra holes of whitespace, the following algorithm was used:

Algorithm 2:

1. Sample $(x_1, y_1), ..., (x_n, y_n)$ from Unif(0, 1),

2. Sample $(c_1, d_1), ..., (c_m, d_m)$ from Unif(0, 1), to be the center of each area of whitespace

3. Given set radius r, remove all points from the uniformly generated data that lies within r Euclidean distance from any point (c_i, d_i)

3.2 Generating Chernoff Faces

The R built-in iris dataset was used to generate images with 3 Chernoff Faces for comparison. In this data, 4 measurements are taken of irises, all of which are categorized as one of 3 varieties: setosa, versicolor and virginica. The measurements, or variables, are sepal length, sepal width, petal length and petal width. In order to generate faces with both subtle and bold differences, as well as faces showing a variety of the possible spectrum of features, each set of 3 faces either contained either 3 different varieties of irises or 3 of the same variety. Four different mappings of facial features to variables were also used to explore which features people may notice more. The faces function in the R package aplpack was used, which maps variables to 15 features total, which are height of face, width of face, structure of face, height of mouth, width of mouth, smiling, height of eyes, width of eyes, height of hair, width of hair, style of hair, height of nose, width of nose, width of ear, height of ear. All plots had the 3 faces in random order, though the same plots were used for all participants. The features were assigned to variables as follows:

1. All features: all 15 possible features of the face are associated with one variable, and the
2. Repeating features, variables are repeated as is default in the faces function (R aplpack faces), excluding the features face shape, height and width
3. Most important features: the 4 seemingly most important features were selected to represent the 4 variables, namely eye height, hair style, smile and ...
4. Least important features: the 4 seemingly least important features were selected to represent the 4 variable, namely face height, ear something something....

3.3 Overview of Plots Used

Overall, 30 plots were shown to participants, with the details in Table 3.1. See algorithm details for parameter definitions.

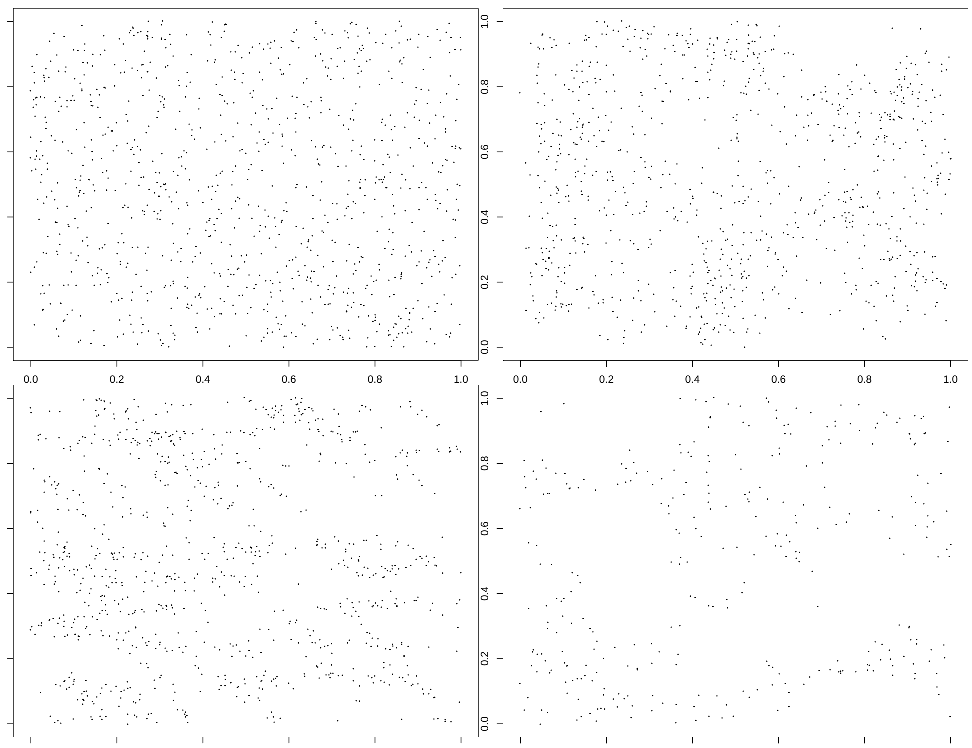

Figure 3.1: Sample scatterplots from slideshow. Top Left: Generated from UNIF$(0, 1)$. Top Right: Data with artificial circular clumps. Bottom Left: Data with artificial lines. Bottom Right: Data with artificial holes of whitespace.

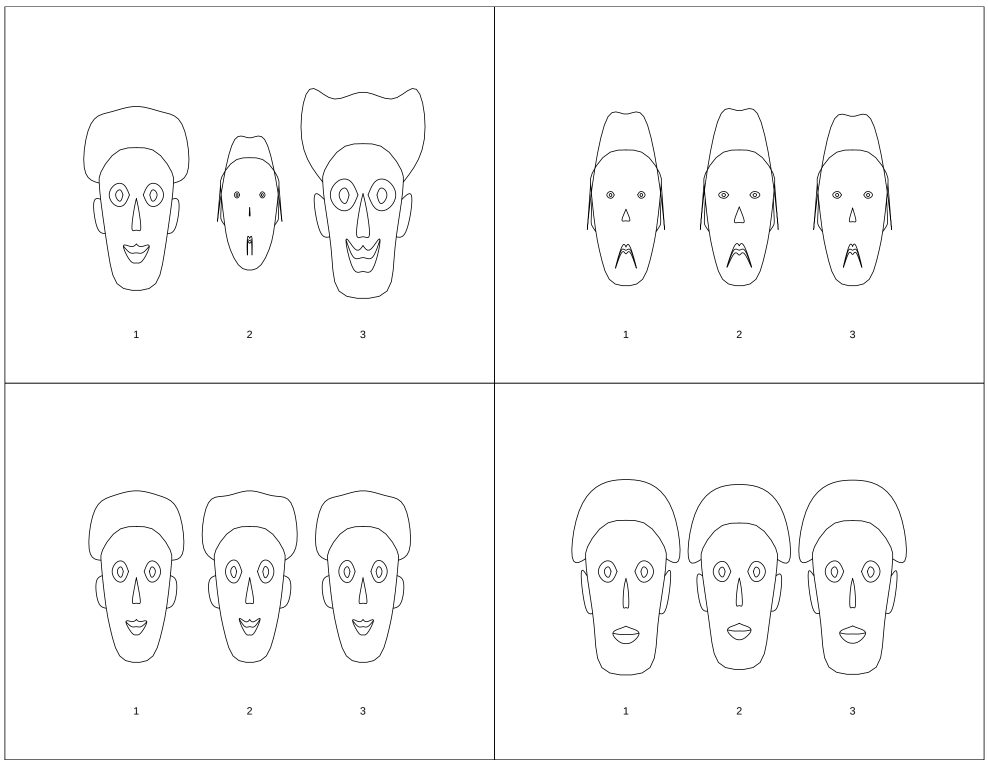

Figure 3.2: Sample chernoff faces from slideshow. Top Left: One iris from each category, one variable, all facial features. Top Right: All setosa irises, all 4 variables, all facial features used except face shape, width, and height. Bottom Left: All versicolor irises, one variable used, facial features used are: smiling, height of eyes, style of hair and width of ear. Bottom Right: All verginica irises, one variable used, facial features used are: height of face, width of face, shape of face and height of ear.

Plot Type	**Plot number**	**Description**
Scatterplots	1-2	Randomly generated from UNIF$(0,1)$, one with 500 points, the other with 1000.
	3	Clumps: $n = 100, m = 900, \sigma_x = \sigma_y = 0.05$, no rotation.
	4	Lines: $n = 100, m = 900, \sigma_x = 0.009, \sigma_y = 0.1, \theta = \pi/2 or \theta = \pi/4$. Lines, half of which are horizontal and half of which are at an angle of $\pi/4$
	5	Lines and Clumps: $n = 100, m = 400, \sigma_x = 0.08 or 0.05, \sigma_y = 0.08, \theta = \theta = \pi/4$. Lines at $\pi/4$ angle and clumps with larger variance.
	6	Lines and Clumps: $n = 100, m = 900, \sigma_x = 0.01 or 0.07, \sigma_y = 0.08, \theta = \theta = \pi/4$. Vertical lines and slightly asymmetrical clumps.
	7-8	Holes: $n = 500, m = 15, r = 0.05, 0.1$. Essentially 15 holes, in one plot with radius of 0.05 and in the other radius of 0.1.
	9-10	Plots taken from real data
Plot Matrix	1-2	Scatterplot Matrix: 20 plots in a grid taken from previous experiments.
	3-4	Boxplot Matrix: 20 boxplots in a grid, taken from previous experiments.
Chernoff Faces	1-4	Only one variable used, repeated over all facial features. One plot using one iris from each type, the others using all from one type for each of the other 3 types.
	5-8	All variables used, each of the 4 is repeated on 3 facial features. Face shape, face height and face width are not included. One plot using one iris from each type, the others using all from one type for each of the other 3 types.
	9-12	Only one variable used, repeated on the 4 features chosen to be most noticeable. The four features are smiling, height of eyes, style of hair and width of ear. One plot using one iris from each type, the others using all from one type for each of the other 3 types.
	13-16	Only one variable used, repeated on the 4 features chosen to be most noticeable. The four features are height of face, width of face, shape of face and height of ear. One plot using one iris from each type, the others using all from one type for each of the other 3 types.

Table 3.1: List of plots in slideshow shown to all participants.

Chapter 4

Experimental Design

4.1 Overview of Experiment

With the goal of demonstrating where people look as they interpret plots, the eye tracker was used to record the gaze path on the screen while a person views a slideshow of images. Three categories of images were used, each testing a different data visualization given an appropriate question.

1. **Scatterplots:** Participants were asked if the given scatterplot looks like it was generated randomly from a uniform distribution. Before the start participants were shown an example including 5 uniformly random plots and one not random plot.

2. **Matrix of scatterplots or boxplots:** Here images from a previous study were used (citation), in which 20 plots are shown as a matrix. One of the 20 plots contains real data, and the others contain data generated to approximately follow the same distribution. There were 2 images with all scatterplots to compare, and 2 images with all boxplots to compare. The participants were asked which plot looks most different from the others, using any criteria they think is appropriate. Before starting, participants were shown a panel of 6 scatterplots and a boxplot in order to explain the question and inform the participants about the basics of boxplots.

3. **Chernoff Faces:** In this section, Chernoff faces generated from the iris data set (default in R) were presented in sets of 3, and participants were asked which face looks most different from the others. Before starting, participants were shown a set

of Chernoff faces generated from the mtcars data set (built-in in R), for a quick explanation of what the faces represent.

All participants viewed the same images, but the order within the category of image was randomized individually for every person.

4.2 Participant Demographics

Participants were recruited through the UW gradstudies mailing list, and reimbursed with a $5 gift card. Of 21 total participants, all were either current grad students or had graduate degrees, and all were in the age range from 20-39. There was one more male than female participants, and participants stated a range from 1 to 10 courses taken in statistics or quantitative methods courses. Participants were also asked for both the number of hours spent browsing the web and number of hours computer gaming in a week. All confidential information has been removed from the data, so participants are numbered from 1 through 21.

4.3 Process for Participants

Every participant was given basic information about the study and asked to meet to perform the experiment. Firstly each person needed to sign a consent form, and fill out the demographics information. The process was then explained, by showing where the eye tracker is and describing what types of images to expect. Each person was given a brief explanation about each type of plot, with examples, and an explanation of the questions they were being asked. They then had the chance to ask questions about what was expected of them during the slideshow. The last element of set-up was adjusting the person and screen to be at good distance and angle for the tracker to work. The tracker was then calibrated to the person, and the slideshow started. During the slideshow, the participant was asked to verbally answer the question associated with each slide, which was then written down. When the slideshow was complete, the participant was handed their gift card and signed a remuneration form associated with it.

4.4 Ethics Considerations

In order to sample human participants, clearance must be given from the University of Waterloo Ethics Clearance Committee. Researchers must complete ethics training via an online tutorial called Course on Research Ethics (CORE). Once trained in ethics, a detailed application is completed prior to all testing, which includes information about recruitment strategy, remuneration, consent forms, email templates, and more.

Chapter 5

Results

Data output from the eyetracker looks like a series of coordinates, one set for each eye. Since the eyetracker runs at 30Hz, there are approximately 30 measurements recorded every second, less when the eye tracker does not pick up the person's eyes for a moment. Demographic information and participants' answers to the questions about the plots were also recorded in association with the recorded path.

5.1 Gaze Point Plots

The simplest way of presenting the data recorded is to plot the path data on top of the image as shown. Recorded locations are plotted with a line connecting the successive points. At times of rapid eye movement the exact location of the gaze is not known, but as the gaze fixes on something there is a clump of points indicating the person was investigating that point further. Since there were 30 plots total and 21 participants, there isn't a good way to look at everything at once, so we present some sample plots which lead to more specific questions about the data.

In viewing the plots of the paths, particularly in the plot matrices, it is interesting to note that some participants have very organized looking paths, with mostly horizontal and vertical lines between fixations. It seems that perhaps those people are using a more methodical approach of looking at each individual plot and comparing to neighbors. In others, the path is very chaotic and the participant seems to cover more space of the plot.

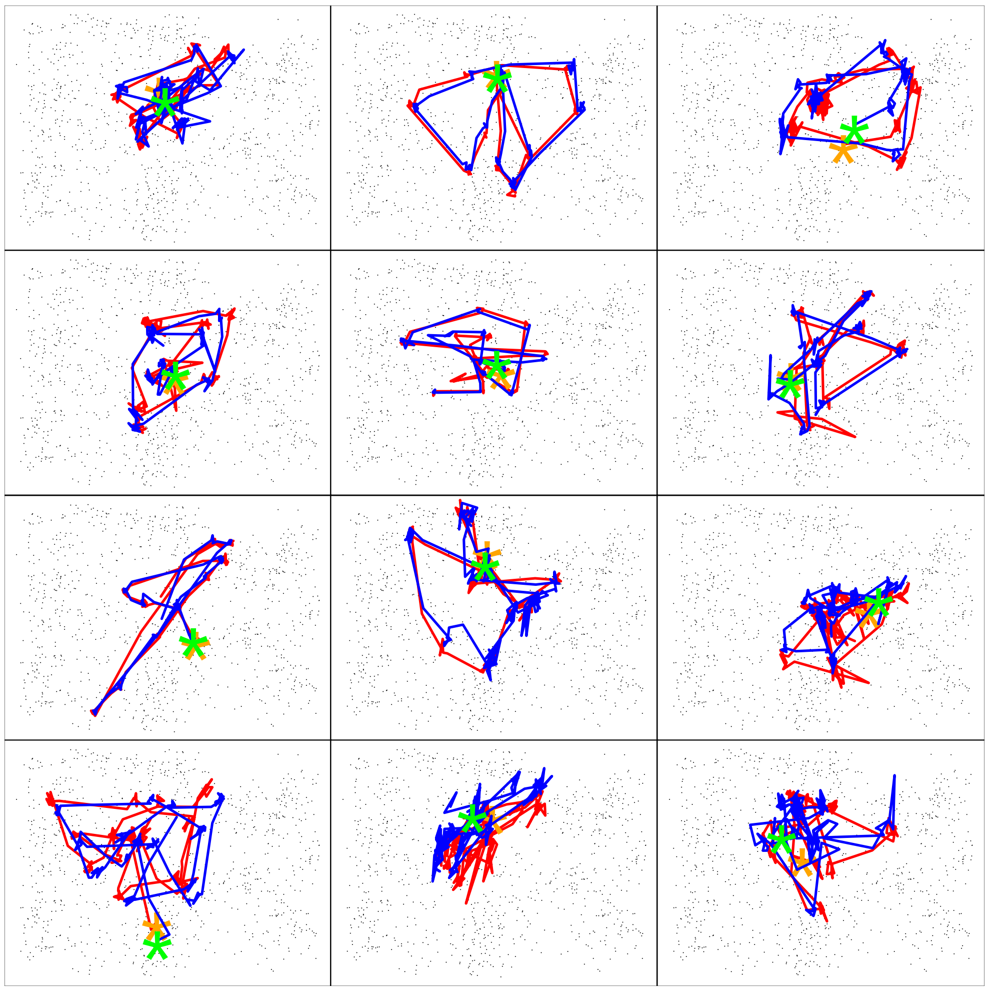

Figure 5.1: 12 participants' tracked eye paths. Red is the left eye, blue is the right. The orange point marks the starting point of the left eye and the green point the start of the right (both of which are off of the screen in some images).

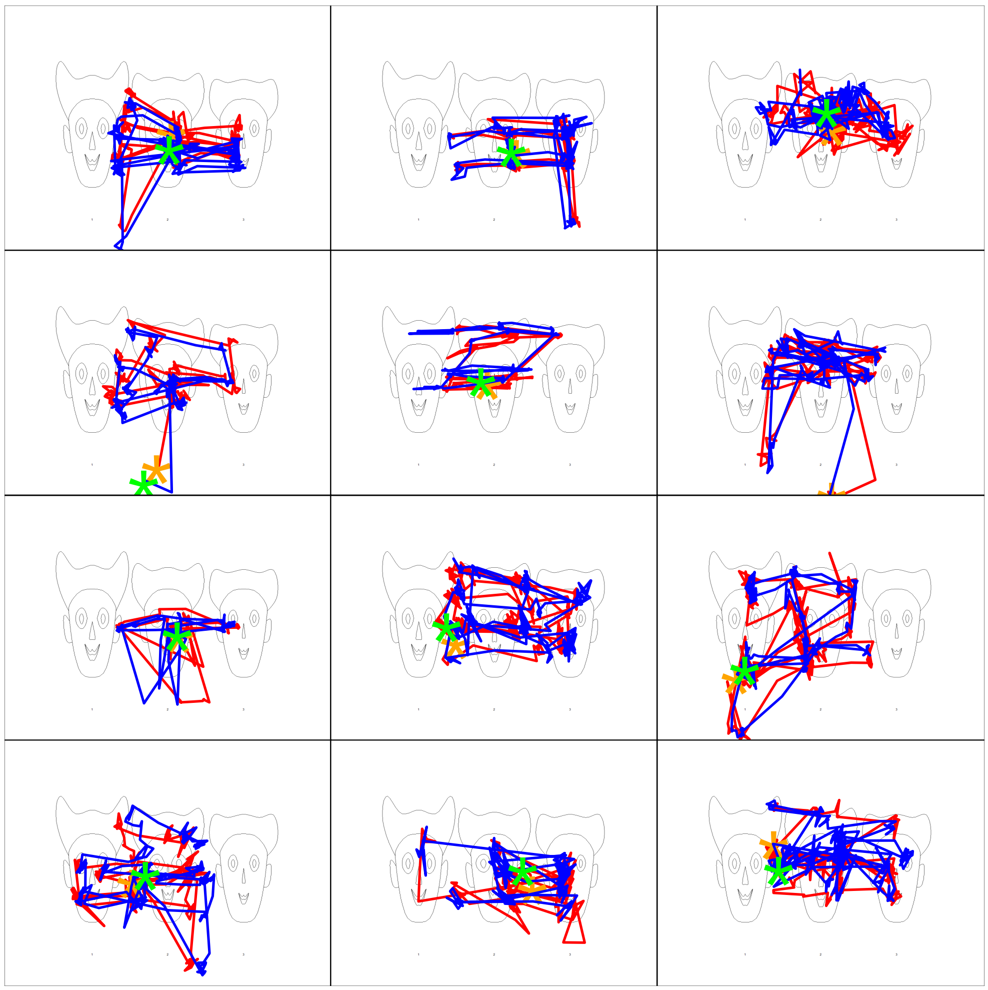

Figure 5.2: 12 participants' tracked eye paths. Red is the left eye, blue is the right. The orange point marks the starting point of the left eye and the green point the start of the right (both of which are off of the screen in some images).

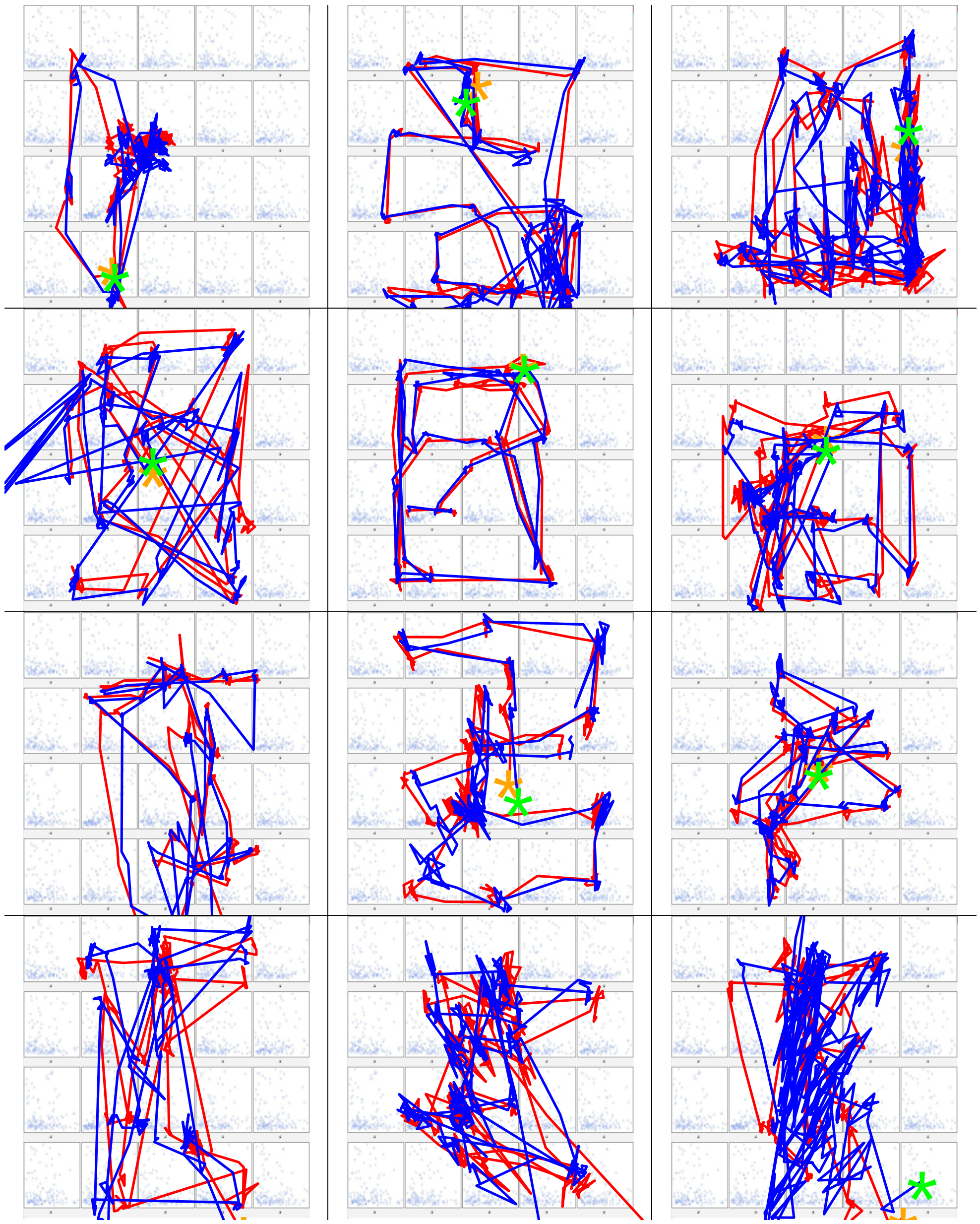

Figure 5.3: 12 participants' tracked eye paths, participants shown here chosen based on the amount of data available. Red is the left eye, blue is the right. The orange point marks the starting point of the left eye and the green point the start of the right (both of which are off of the screen in some images).

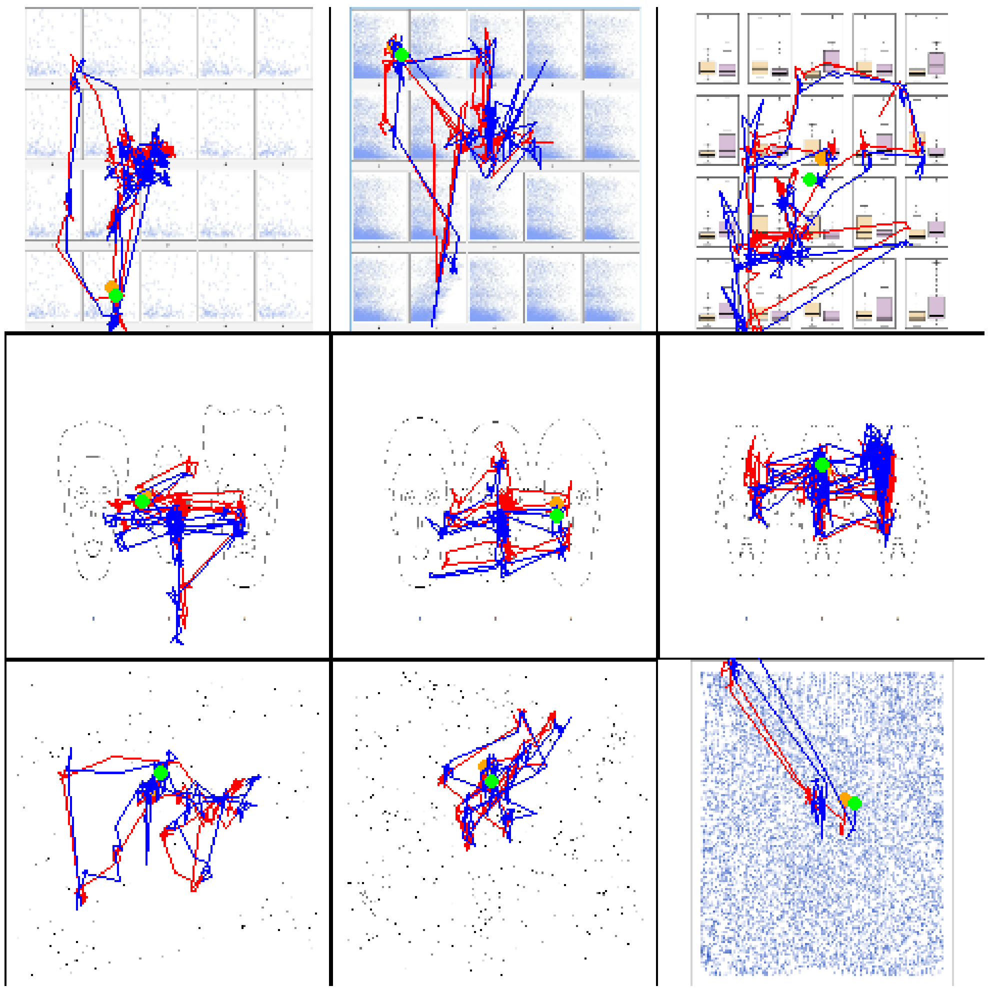

Figure 5.4: Participant 2, eye path on a subset of the images shown. (Terrible quality to be fixed here)

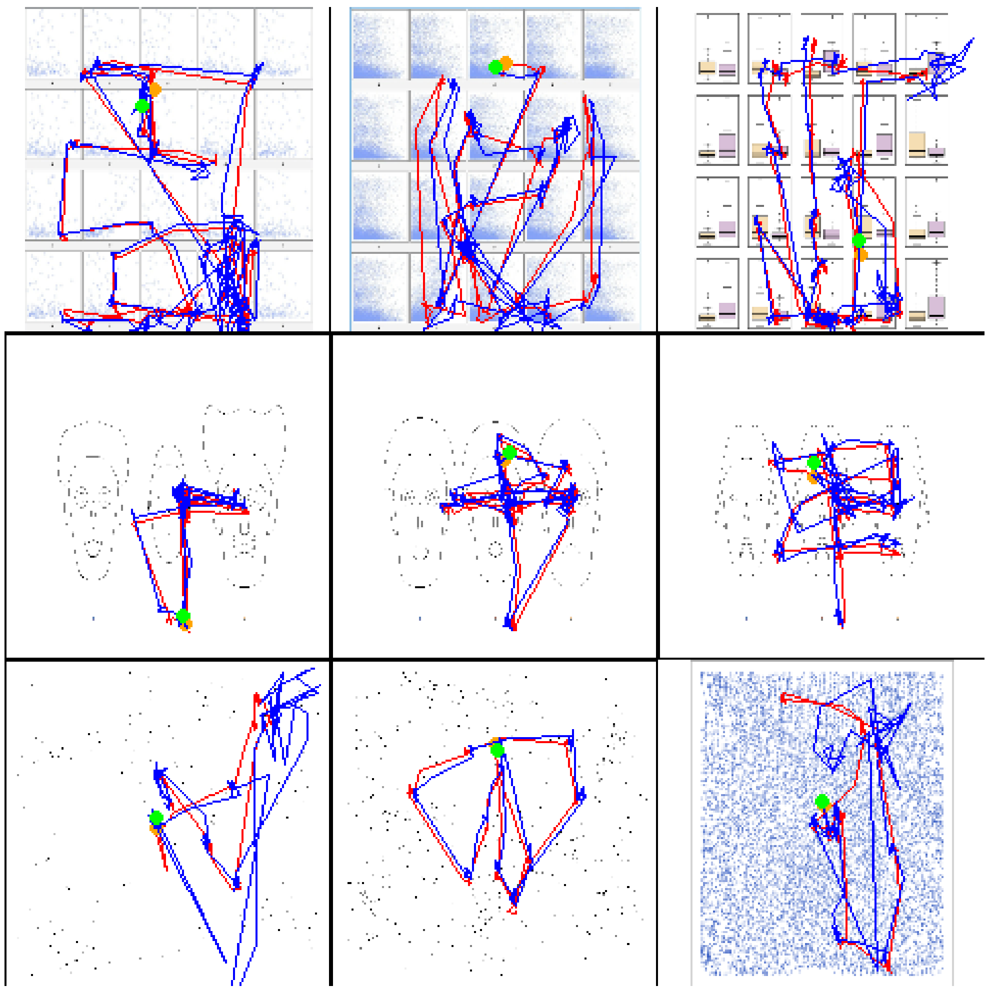

Figure 5.5: Participant 3, eye path on the same subset of images as previously. (again, quality to be fixed)

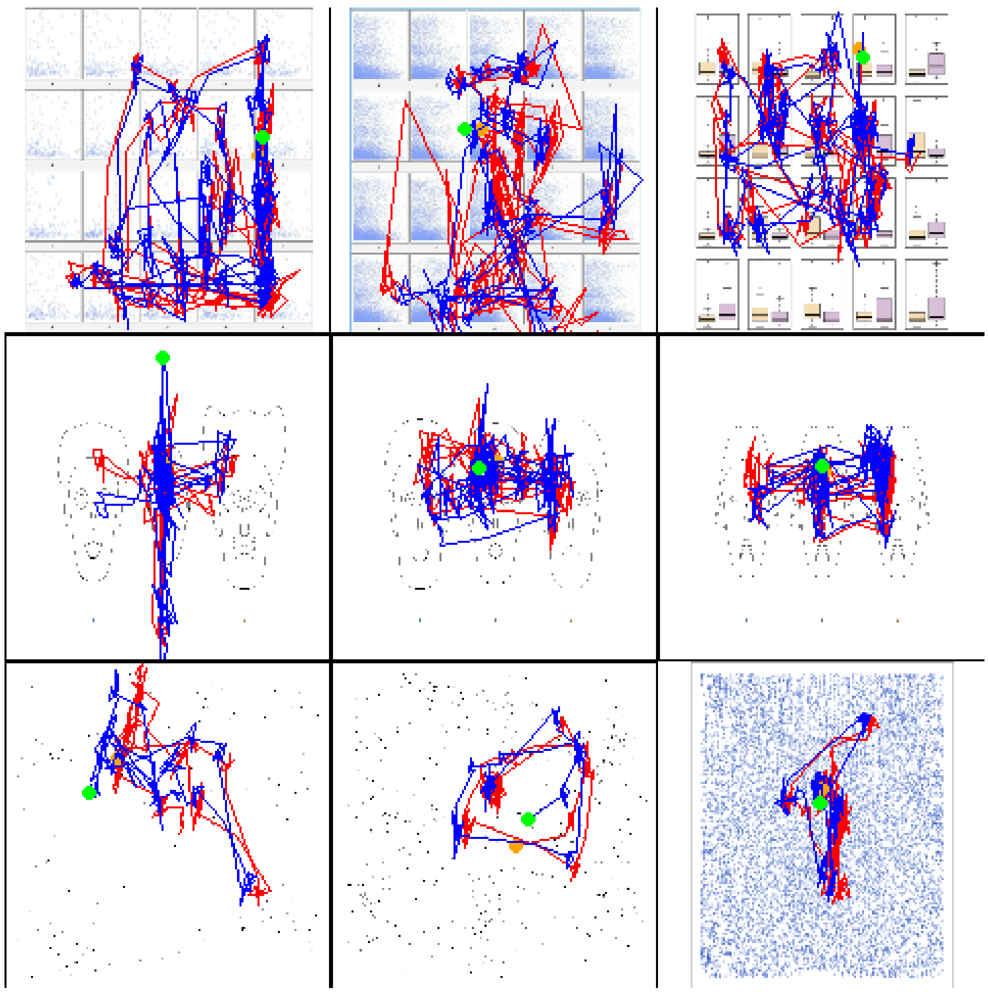

Figure 5.6: Participant 7, eye path on the same subset of images.

5.2 Participant Comments and Inconsistencies in Data

While most participants had no trouble finding a position where their eyes were well tracked, of the 21 one person was not consistently tracked no matter where they were sitting. As a result, for that person not every image shown has an associated path, and the paths for images that were recorded are probably less useful since there are large gaps in time between the tracker following the participant's eyes. Although in the testing stages eyeglasses appeared to not affect the ability of the tracker, for two participants the tracking was working better without glasses, and those participants chose to remove their glasses for the duration of the tracking. Other causes of missing data may be from the participant blinking, the participant gesturing and blocking the light from the tracker, or the participant moving out of range during the slideshow.

Since the slideshow ran continuously with no way to pause or interrupt, all participants had the same length of time viewing each image, but some found it challenging to provide an answer within the time. From those participants who found it most difficult, I heard feedback after that it was frustrating that they did not have time to look at everything before deciding. Since many other participants didn't have a problem, it seems that people were using different approaches to answer the questions, and some people did not have time to produce a confident answer. In the future, this may be an area to address with either an interactive slideshow with the participant controlling it, or with a sample slideshow to start out, so that participants know what the given amount of time will be and can adjust their methods.

Chapter 6

Conclusions

To conclude, the experiment described here was a success given that there was data collected on the path of participants' gaze when observing statistical graphics. The Tobii eye tracker used was portable and provided good quality of data for a variety of users. Participants in the study were asked to view a slideshow of plots and answer simple questions, for which their eye were tracked and answers were recorded. All participants were able to answer the questions and did not have trouble understanding the process. All the plots used in the study are reproducible and all scatterplots and Chernoff faces were generated methodically in order to cover a variety of cases.

Missing data from the tracking is a concern that is not easy to fix given the current system. One person seemed to not track as well as the others, and even with participants with a consistent track there are inconsistencies due to blinking, gesturing or moving out of range. The missing points would be more of a concern for a more thorough analysis, so for the purposes of this report only the participants with the best data (most gaze points tracked) have been displayed.

In the future, further analysis of the current data would be recommended since this report does not use any quantitative methods of drawing conclusions based on the data. It would be interesting to look at the relationship between the participant's answers and how their eye tracking plots look. Clustering algorithms could be applied to determine if there are similarities between people's response based on their path. This study was a fairly quick first look at where people look when they view plots, and using this as a starting point there could be countless other investigations involving different styles of plots, different questions, a more interactive system of testing, and of course further analysis.

References

[1] Alfred L. Yarbus. Eye Movements and Vision. Plenum Press, New York, 1967.

[2] Steven Pinker. Visual cognition: An introduction. Cognition, 18 (1-3): 1-63, 1984.

[3] Patrick Cavanaugh. Visual Cognition. Vision Research, 51 (13): 1538-51, 2011.