Replication of “Regression by Eye: Estimating Trends in Bivariate Visualizations” by Michael Correll and Jeffrey Heer

Bianca Stancu, Alix d’Agostino, Natalia Obukhova, Tim Brlan, Alexander Pfyffer, Mathias Lüthi

20.05.2019

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Motivation

Our research intends to verify if the results of the "Regression by Eye: Estimating Trends in Bivariate Visualizations" by Michael Correll and Jeffrey Heer study can be applied to other participants, setup, software, and data without statistically significant deviation. Therefore, we will need to check how different data visualizations and biases affect participants' estimation of trends using regression by eye by performing replication of their study. We have decided to conduct the first out of the three experiments from the original study, because we found it to be paramount to the main topic of the entire paper since it serves as the baseline for the next two experiments. Furthermore, as the second and third experiments simply added to the first by including bar charts and outliers respectively, we found the first experiment to be the most compelling to replicate.

1.2. Related Work

The original study [1] concluded, that viewers without statistical training accurately estimated trends in many standard visualizations of bivariate data. Specifically, "larger residuals result in less accuracy at regression by eye" [1]. Moreover, there was no statistically significant difference in estimation accuracy among linear, quadratic, or trigonometric trends. Finally, there was no statistically significant bias in estimations.

1.3. Research Question

Our main research question is "How do the type of chart (scatter plot, line graph, area chart), the bandwidth of the residuals and trend type (linear, quadratic, trigonometric) affect the estimated trend of the bivariate visualization by an individual?".

1.4. Hypotheses

For the experiment, we will use the confirmed hypotheses from the original study and use the negations of the failed ones as our own. The first hypothesis is that participants will see a balanced set of positive and negative trends with no bias in estimations from under- or overshooting (H1), as we do not expect that the participants will have a penchant for either trend. The second suggests that larger residuals will result in less accuracy in participants' estimation (H2), because as the residuals increase, the data on the graph will appear more scattered and less precise, making it harder to estimate accurately. Finally, our third hypothesis is that the type of trend will have no statistically significant effect on participants' estimation (H3), as we do not expect the participants to be hindered by the complexity.

1.5. Key Factors

For our key factors, we will contrast three different variables that are chart types, slopes and the bandwidths of residuals. Our first independent variable is the chart type, which has three levels: scatter plot, line graph and area chart. The second will follow with the slope for the linear fit, the curvature for the quadratic fit, "controlled by the coefficient of the second term" [1], "the amplitude of the cosine function" [1] for the trigonometric fit, including eight different levels (-0.8, -0.4, -0.2, -0.1, 0.1, 0.2, 0.4, 0.8)Lastly, the bandwidth of Gaussian residuals will be our third independent variable with four levels (0.05, 0.1, 0.15, 0.2). Also**,** we use an additional random factor - the chart type with 3 levels (area chart, scatterplot, line graph).

1.6. Operationalization of the Measurements

We will present the participants with a series of bivariate visualizations, who will adjust a slider to fit the perceived trend. The slider parameterizes one of three types of the trend: linear (the slope), quadratic (the curvature for the quadratic fit, "controlled by the coefficient of the second term" [1])), or trigonometric ("the amplitude of the cosine function" [1]) in range [-1, 1]. The functions go as follows:

- Trigonometric: $y = 0.5 - slider_{value} * 0.5 * \cos(x * \pi)$;
- Quadratic: $y = 0.5 + slider_{value} * x^2 - 0.5 * slider_{value}$;
- Linear: $y = 0.5 + slider_{value} * (x - 0.5)$;

where $slider_{value}$(the slope) is controlled by the slider in the data collection software in range [-1,1]. For each stimulus, the participants will adjust a slider that controls the slope of a rendered trend line. In the case of quadratic trends, this slope is the curvature; for trigonometric fits, the positive/negative amplitude.

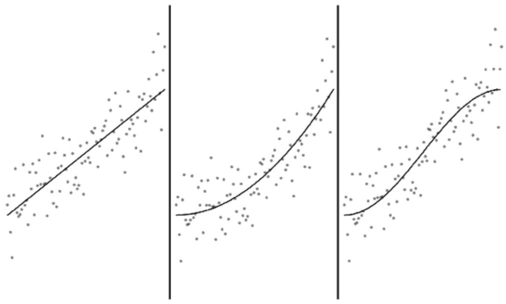

Figure 1 - The three different trend types in this study: linear, quadratic, and trigonometric trends. Source: [1]

Besides our main results, we will also collect the following demographic information from the participants as potential covariates:

- Age;
- Gender;
- Educational level (e.g., High School, BSc/BA, MSc/MA);
- Field of Study;
- Nationality;
- Familiarity with charts and graphs with an emphasis on scatter plots (tested on a 5-point Likert scale).

We decided to add the nationality in the demographics, as the original experiment was performed with subjects coming strictly from the United States. If we obtain different results during the replication, this factor could be causal. Just like in the original study, the familiarity with charts and graphs is relevant as it can explain the fact, if it does occur, that some participants perform better than others because of their experience.

1.7. Predictions

To commence, we predict that participants require an adaptation period to become accustomed to the interface of our research software, as well as to the hardware used during the experiment. To remove this adaptation time from the final results, we added three practice trials. After which, we expect our experiment sessions to yield close results to the original study.

1.8. Key Risks

Below is a list of potential risks that could be involved in the experiment process and have analyzed methods to decrease the probability of them occurring. First, if a participant lacks a basic understanding of graphs due to insufficient exposure, this would negatively bias our results. Likewise, this can be applied to a basic knowledge of statistics, where individuals with a superior statistical background could be favored. We intend to record this knowledge in the demographics form (see Appendix 9.2), to verify that it does not create a bias.

Since the participants are completing almost a hundred tasks that are somewhat similar, fatigue could be an issue if participants are not accustomed to working on a computer. This led to including a criterium regarding computer use, which must be more often than once a week. If participants do not engage in step estimation according to our specific tests, it would cause errors through the combination of estimation techniques. We expect this risk to be reduced to a minimum by having clear steps and directions, while explicitly explaining them to the participant before the study. Finally, the software we have created could have some bugs or even potentially crash. We plan to conduct four pilot studies and debugging sessions in order to mitigate the risk of it occurring.

2. PARTICIPANTS

2.1. Participants Factors

In this study, we expect to find differences between participants who regularly view, analyze or produce charts and/or graphs to be better at estimating regressions, compared to those who are not often exposed to these. This will be covered during part of the briefing before the experiment to establish if as anticipated this factor has an influence. We believe this factor is likewise indirectly linked to the participant's studies, which will be added in the participant demographics. Participants studying psychology, for example, are more likely to be exposed to graphs than a participant studying philosophy. Age and degree level are also likely to have a slight influence, as older participants or those that have studied longer are more probable to have been exposed to graphs and therefore could have garnered experience in graph reading and interpreting. We do not expect to find significant differences between genders; however, we intend to include a balanced ratio of genders in the participant criteria. Additionally, we decided to ask the participant's nationality, to ensure that it does not influence the results. We do not expect it to have an effect. Lastly, we will inquire in the participant's past experiences with experiments. We believe that participants who have done many experiments, especially if they were in the HCI field, could have an advantage at estimating the regressions.

2.2. Participant Criteria

For a successful study conduction, we will need a multiple of 8 participants. In order to get the statistical power above 0.8, we will need to hire at least 16 participants (see fig.2). To exclude gender as a potential bias, we will have approximately 50-50 ratio of men and women in our experiment. All the participants must be 18 years old or older in order to avoid including minors in the study, who are considered to be a vulnerable population. Additionally, the participants cannot have significant vision problems, as that would prohibit them from performing well compared to others, not due to their regression estimation skills, but rather their inability to clearly see the graphs. Lastly, the participants must be familiar with computers, as we believe that individuals who do not use computers on a regular basis are likely to perform poorer due to their inexperience with the platform. Finally, for obvious reasons, individuals who know the study we are replicating cannot participate in the experiment as it would create a significant bias. No other requirements apply.

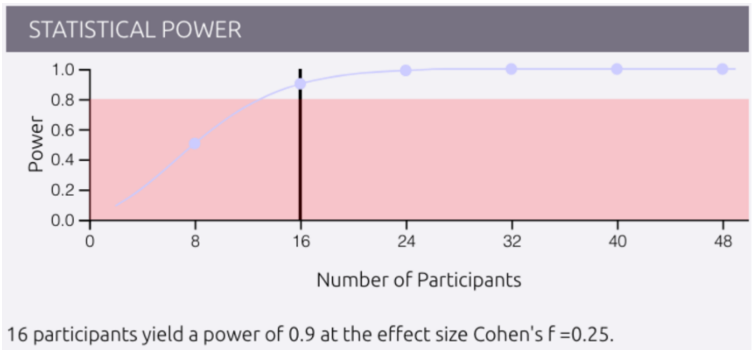

Figure 2 - Statistical power

2.3. Recruitment Strategy

As explained in the participant criteria, we wish to find participants who are comfortable with computers. In order to do so, we assessed that we would have the highest rate of success by recruiting in university buildings, such as BIN in Oerlikon. We plan to entice students through chocolate compensation as motivation for participating, as well as the short time necessary to complete it. We also hope the individuals will be motivated by doing a good deed (research karma) and furthering the knowledge on data visualization. In order to increase efficiency, we intend to have one to two researchers searching for new participants, while the others are conducting the experiment.

2.4. Compensation

We intend to offer chocolates to encourage people to participate in our experiment and to compensate for their time and effort. As our experiment is slightly time-consuming, we believe that chocolates are motivating enough for potential participants. Drinks and snacks will also be available in the briefing and debriefing rooms to make the participants more comfortable.

2.5. Debriefing

After the experiment has taken place, we will ask how the experiment went and note down any comments they wish to express. Then, we thank the participant for their effort and give them a bar of chocolate as compensation. If they wish to increase their knowledge about the experiment design and/or the question at hand, we will inform them accordingly.

2.6. Risks & Discomforts

There are no risks, physical or psychological, linked with the participation in this experiment.

Some discomforts or fatigue could be felt for participants that are not accustomed to computer screens since they will have to perform by staring at a screen. We expect to eliminate this possible discomfort by only recruiting participants that use a computer more than once a week, to whom this should not apply. Nevertheless, if a participant should feel uncomfortable, a member of the research team is always available to help them. The experiment can be paused (if the participant is willing to continue) or fully stopped (if the participant is not willing to continue). The participants are informed of these possible discomforts before the experiment begins during the informed consent part.

2.7. Benefits

Since the participants in this experiment will have to estimate regressions by eye, their skills in graph reading could somewhat improve. Participants will, therefore, get a slight improvement in estimating trends and regression lines. Also, they will learn some basics on graphs and regressions during the briefing before and after the experiment, which is a useful skill to have. Finally, and most importantly, the participant will allow us to grow our understanding of regression by eye and how the type of chart, slope and bandwidth influences an individual's interpretation of data.

3. APPARATUS (HARDWARE AND SOFTWARE)

Software: We will be using our custom interactive software, where participants will be presented with different plots and their corresponding sliders to adjust the trend line. The functions are described in 1.6. Once the software is launched, a blank page with a "choose participant" drop down list will appear. One the participant ID has been selected, the welcome screen appears. "Welcome to our study! Thank you for the participation in our experiment! Please listen to the instructions and, once instructed, click on the button below to start practice tasks.", below a "Start the practice" button will appear. The participants must click to start the three practice trials (see 4.1 for details on said trials). Then a graph appears with the programmed setup of the variables and trends. The participant will have to use the slider below the graph to change the slope of the trend line, ranging from -1 to 1. Once the participant has chosen the slope, they click the "Next" button below the slider to advance to the next stimulus. If the participant clicked "Next" accidentally, they may click the "Previous" button to go to the previous graph, avoiding this risk completely. Once the practice trial is over, a new blank page with a "Begin experiment" button will appear. Once ready, the participant will go through the 4 validation and 96 main tasks. Once over, a blank screen will appear with "Thank you for participating in the experiment. Please go back to the introduction room for you debriefing and compensation".

The software is a web program coded in JS, HTML and CSS, and can be found in our Github repertoire[1]. The software takes as input a CSV file with the trial tables and produces as output the same trial table with the participants' answers added accordingly.

Hardware: The participants will be using desktop computers with mice, running our software. We have chosen to use a mouse during the study, as trackpads on laptops vary greatly between brands and models, which could negatively affect results due to difficulties with the handling of the trackpad. This will allow the participant to be more comfortable and accurate during the trials.

Setting: Our study will require two rooms: One for the participant briefing and debriefing, and another for the experiment. The experiment room will be equipped with several workstations for the experiment. This will allow us to conduct the experiment with several participants simultaneously, where each participant can complete the experiment in

[1] https://github.com/mathiasluethi/ReplicationExperiment-RegressionByEye

their own time. Participants will be recruited by one researcher and accompanied to the first room, where they will be briefed by another researcher and led to the experiment room. After finishing, the participant will return to the first room for the debriefing and to receive the compensation.

4. PROCEDURE (PRECISE AND REPLICABLE)

4.1. Experimental Design

The goal of this experiment is to examine how accurate participants are at estimating the magnitude (slope, amplitude or curvature) of trends in bivariate visualizations. To measure this, we will conduct a study, where participants have to estimate trendlines for generated data visualizations based on predefined conditions.

Each participant will view a combination of three chart types, eight possible slopes and four bandwidths of Gaussian residuals, for a total of 100 (96 + 4 validations) stimuli. Each participant will perform three additional practice trials at the beginning of the session that will be excluded from the analysis. The data will be pre-generated for each participant.

These give us our first set of *independent variables:*

- Chart type with three levels: scatter plot, line graph and area chart;
- Slope with eight levels: $\beta = \{-0.8, -0.4, -0.2, -0.1, 0.1, 0.2, 0.4, 0.8\}$;
- Bandwidth of Gaussian residuals with four levels: $\sigma = \{0.05, 0.1, 0.15, 0.2\}$.

These independent variables will be tested through a within-subject design, meaning that each participant will see a visualization of every combination of these three independent variables. The chart type variable will be randomized, however remaining in a block so the participant does not have to switch content between every stimulus. The remaining two variables follow the Latin square, each in their own block.

There are three types of trends (linear, quadratic, trigonometric) used and each type is randomly assigned 32 times to a stimulus (see Appendix 9.3 - Trial table). The slope of each trend type will be adjusted differently with the slider. As assigned, it will evenly distribute the trend types between the 96 stimuli.

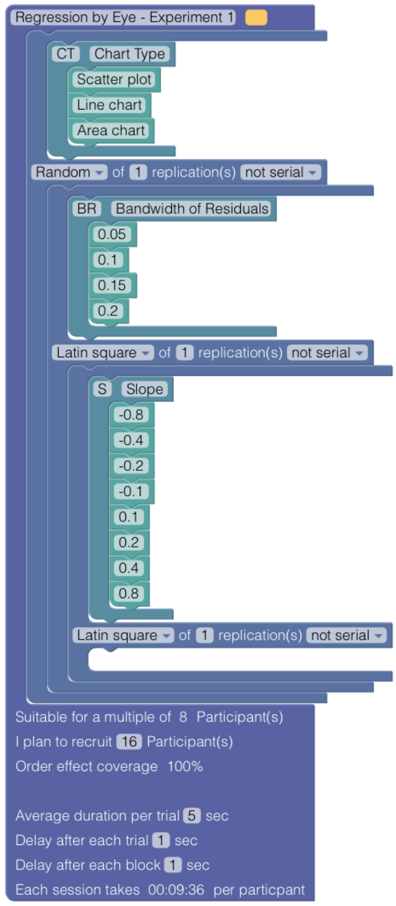

Figure 3 – Experimental design

While the trendline is linear the slider will control the slope of the rendered trendline. In the case of the quadratic trends, the slider will change the curvature and for trigonometric fits, it will change the positive/negative amplitude of the trendline. We will also include four validation stimuli for each participant (the trial, where the correct answer is explicitly provided, i.e. the correct trend line is shown on the chart), but these data points will however be excluded from the final result.

4.2. Blocking

To keep interruptions and switching tasks at a minimum, all participants will get a general introduction and then work through all tasks without any significant breaks. We won't impose a time limit, so each participant can work at their own pace. We decided to make three blocks based on the chart type so the participants will not be interrupted by the change of context every task.

4.3. Counterbalancing

All the independent variables will be combined to create 96 different tasks (8 slope levels * 3 chart types * 4 bandwidths, which gives 96 for all possible combinations) plus 4 validation tasks (with randomly chosen order and variables) for each participant to complete. We chose random counterbalancing strategy for the chart type and the latin square for bandwidth and the slope. Also, the additional factor added randomly assigned 32 times of each type of the three trend types (see Appendix 9.3).

The participants counterbalancing table is included in the Appendix 9.3 at the end of this document.

4.4. Training / Practice

We are giving the participants a set of instructions on their tasks and how to use the study's software. Before the start of the experiment, they will be asked to complete four trial estimations without recording the data in order to exclude any potential misunderstandings.

For the practice tasks, one of each trend type and chart type will be performed, therefore, variables have been chosen to be the same for all participants and are as follows:

- Scatter plot, slope = -0.4, Gaussian residuals σ = 0.15, trigonometric trend;
- Line graph, slope = 0.8, Gaussian residuals σ = 0.1, quadratic trend;
- Area chart, slope = 0.2, Gaussian residuals σ = 0.2, linear trend.

4.5. Conditions

Each participant will see all the combinations of the previously specified visualization of stimuli properties:

Chart type (CT):

- CT1: Scatter plot
- CT2: Line graph
- CT3: Area chart

Slope (S):

- S1: -0.8
- S2: -0.4
- S3: -0.2
- S4: -0.1
- S5: 0.1
- S6: 0.2
- S7: 0.4
- S8: 0.8

The bandwidth of Gaussian residuals (BW):

- BW1: 0.05
- BW2: 0.1
- BW3: 0.15
- BW4: 0.2

To each of the conditions above, the previously specified trendline properties (TL) will be assigned:

- TL1: Linear trendline
- TL2: Quadratic trendline
- TL3: Trigonometric trendline

4.6. Tasks

Participants are presented with a series of bivariate visualizations. The visualizations consist of a graph and a trendline. For each stimulus, participants have to adjust the trendline with the help of a slider to match the perceived trend.

4.7. Step-by-Step Description / Timeline

1	Introduction	30 seconds
2	The consent form, demographic form & questions	5 minutes
3	Step-by-step instructions for using the software	45 seconds
4	Completing practice tasks (4-10 seconds * 3 tasks)	12-30 seconds
5	Main task begin & questions from the participant	30-60 seconds
6	Completing the main tasks (3-10 seconds * 100 tasks)	5-17 minutes
7	Final debriefing, questions & compensations	30-60 seconds
8	*Buffer time (movement & unexpected)*	*2 minutes*

We, therefore, estimate the experiment to take on average approximately 17 minutes per participant, ranging from 14 to 28 minutes.

4.8. Experiment Summary

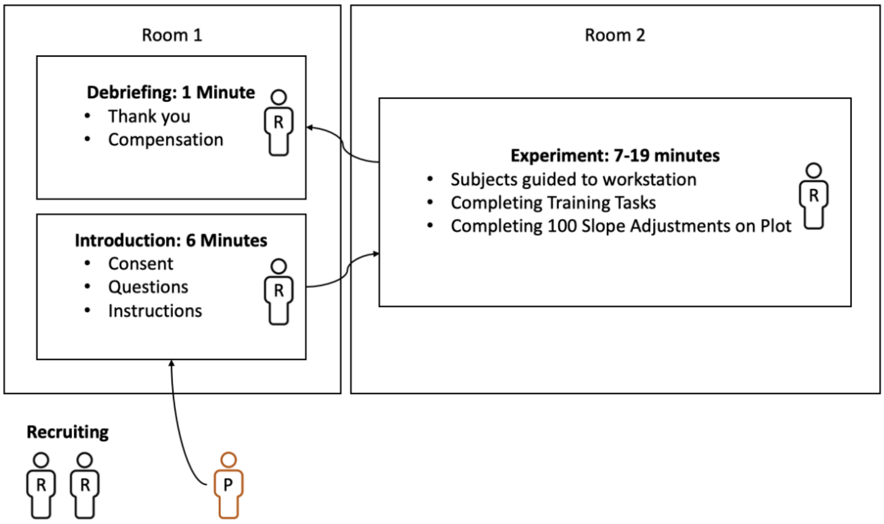

Figure 4 - The experimental process. “R”s are researchers, “P” is the participant.

5. DATA COLLECTION & ANALYSIS

5.1. Data Collection

Quantitative:

- The bandwidth of residuals (discrete);
- Slope;
- Chart type (three categories: Area, Line graph, Scatterplot);
- Trend Type (three categories: Trigonometric, Quadratic, Linear).

Qualitative:

- *No data.*

The data will be collected automatically, during the experimental procedure by the software designed for this purpose. In order to generate the points used in the collection, the *Processing* script[2] from the original study will be used.

5.2. Unit of Analysis

One trial is defined by one instance of all the previously mentioned variables. It corresponds to the participant performing one estimation of the regression given a set of points, using the available input controls. All 96 trials with the additional 4 validation trials will be performed in three blocks, based on the chart type, in one session. The validation data points will be excluded from the analysis.

5.3. Formatting & Filtering

In order to restrict the possibility of data corruption, on the spot validations will be performed. The data from participants with a high average error (more than 0.2) on the validation tasks will be excluded from the analysis. After the check on validations, the data collected on these tasks will be excluded from the analysis.

5.4. Error Management

The following participant errors can occur:

- Software crashes and mistakes in data collections: all the csv files will be checked for the completeness
- Submitting the answer by mistake is handled by adding the button "Previous" in the software

5.5. Proposed Data Analysis Strategy

The first step is the post-processing of the data, which includes data cleaning (missing values, anomalies; see 5.4) and computation of the error in estimation of trend line, as the absolute difference between the two slopes. The error in estimation will be calculated as follows:

$$error = (the\ slope\ from\ the\ OLS * sign\ of\ the\ slope) - the\ slope\ from\ the\ participant$$

The slope in the equation above is considered as such: the *slope* for the linear fit, the *curvature* for the quadratic fit, "controlled by the coefficient of the second term" [1], "the *amplitude* of the cosine function" [1] for the trigonometric fit.

The resulting format of the data, post-processing, will match the format from the original experiment. The output of this step will be a csv file that will be used as input to the data analysis step, with the following columns:

- index: the number of the task from 0 to 99;
- sigma: the residual bandwidth;
- sign: sign of the slope;
- type: chart type (line, area, scatter);
- m: absolute value of the slope of the trend line;
- id: participant identifier;
- graphtype: trend type (line, quad, trig);
- answer: participant's answer;
- isValidation: empty or 1 if the task is a validation task;
- error: difference between slope and answer (m*sign-answer);
- unsignedError: absolute value of the error;
- src: the file path to the image (visualization).

Once the errors have been calculated, the next step will be the data analysis. The process will be entirely done in *R*. The first data analysis step will be to perform a three-way analysis of covariance (ANCOVA), to see if the three independent

[2] https://github.com/uwdata/trend-bias/blob/master/Data/MonteCarloPoints.pde

variables are causal and not due to chance. If they are causal, it will show how they affect the outcome. The suggested approach ANCOVA will be as follows:

- *Independent variables:* residual bandwidth, graph type, and trend type;
- *Dependent variable:* error in estimation of trend lines;
- *Covariates:* participant ids and actual slope of the trend line.

Then, we will test our three hypotheses. With the results we will be able to confirm or reject the null hypotheses and compare our results to the original study. If any of our results are different, we will analyze how they differ and see how we can find explanations for these differences.

For the third hypothesis, further tests to identify pairwise interactions (using *Tukey's Honest Significant Difference*) will be performed and interpreted. For the first hypothesis, a *Student T Test*, with the null hypothesis of *mu(error) = 0* is performed and interpreted in order to check if the positive and negative trends are balanced. In this case, the signed error will be used.

For the ease of the data analysis, we will also visualize the data, using the two plots, as in the original study:

1. x - the absolute error, y - the bandwidth of the residuals;
2. x - the absolute error, y - the trend type;
3. x - participant ID, y - the signed error.

These plots will allow us to directly compare our results to the original study and see how our experiment, if it does, differs. The third plot was not included in the original study, yet we choose to include it to allow a visual representation of the results for hypothesis 3.

6. DATA MANAGEMENT

6.1. Data Storage

The data will be stored in CSV format, one per participant, on a private space shared only with the researchers. Similar to the original study, a CSV file containing demographic information about the participants will be stored as well. The signed consent forms will be stored on the paper. All the data will be purged after 3 years from collection.

6.2. Anonymization

The data will be anonymized from its source, the only collected data referring to the participant's identity being the consent. Participants will be referred to using an ID during the study and in the demographic forms.

6.3. Attribution

No outsourcing initiatives for any work in the context of this study will be undertaken. The only external source used within this study is the points generation script from the original study, which is publicly available without restrictions. Also, we thank ZPAC lab for the Touchstone2 software.

7. BIBLIOGRAPHY

1. Correl, M. and Heer, J. (2017). Regression by Eye: Estimating Trends in Bivariate Visualizations. In: *CHI 2017.*

8. ADDITIONAL DOCUMENT CHECKLIST

8.1. Human Subjects Committee Checklist

Out of the 13 questions in the OEC Human Subjects Committee Checklist, none have been answered "yes" or "don't know", therefore no submission needs to be made. The questions are the following:

- Does the study involve subjects that are uninformed about their participation in the study? – *No.*
- Does the study involve subjects from populations that are vulnerable? – *No.*

- Are the subjects deceived or mislead by the researchers? – *No.*
- Does the study involve coercive financial or non-financial incentives? – *No.*
- Is there a substantial dependency relationship between the subjects and any of the involved researchers? – *No.*
- Could the study negatively influence the subjects' psychological integrity? – *No.*
- Could the study negatively influence the subjects' physical integrity? – *No.*
- Are subjects asked to provide sensitive personal information? – *No.*
- Is the purpose of the study to significantly influence people's lives or real-life? – *No.*
- Does the study expose any of the members of its research team to threats harming their physical or psychological integrity? – *No.*
- Does the study involve the collection of data from voice, image or video recording? – *No.*
- Will the study collect and use data that is not anonymized? – *No.*
- Do any members of the research team have any association that poses or could be perceived as posing a conflict of interest in connection with the result of the study? – *No.*

8.2. Recruitment Emails/Flyers

We do not intend to use neither emails nor flyers. We intend on recruiting potential participants by approaching them and asking if they are willing to perform in our experiment.

8.3. Informed Consent

The participants will receive the following information regarding the procedure and their rights:

- The participants are allowed to end the experiment and leave at any time – without reason or explanation.
- There are no risks associated with this experiment.
- There is the possibility to feel slight discomfort or fatigue from looking intensely at the computer screen.
- All results are anonymized, and no confidential data are kept.
- The description of the procedure.

The full version of the consent form is shown in Appendix 9.1. All the data in the consent forms will only be stored in the signed paper form and will be purged 3 years after the end of the experiment.

8.4. Media Release Form

We allow the reproduction of our experiment for non-profit generating purposes.

8.5. Participant Counterbalancing Table

See: https://tinyurl.com/y5bpe8oq

8.6. Initial Debriefing

Each of the participants will be asked the following questions during recruitment to check if they qualify for the study:

- How old are you? *Must be at least 18.*
- Do you have a visual impairment which would prohibit you from seeing detailed graphs? *Must answer no.*
- Do you use a computer more than once a week? *Must answer yes.*
- Have you read the paper "Regression by Eye: Estimating Trends in Bivariate Visualizations" by Michael Correll & Jeffrey Heer? *Must answer no.*

Also, we will give a short description of the study purpose, data collection, anonymization process and risks and discomforts of the study (see 7.7) and collect the consent forms (see Appendix 9.1). All the additional demographic questions will also be asked before the main experiment via the demographic questionnaire form (see Appendix 9.2).

8.7. Experimenter Instructions/Script

<Participant arrives after passing selection process done by recruiters>

"Welcome to our study! My name is *<say name>* and I will be briefing you on this experiment. I will begin by reading your rights, then explain the procedure of the experiment, then you will do the experiment, and finally, you will receive a chocolaty compensation. Please do not hesitate to interrupt me if you have any questions.

Let's begin by going through your rights and consent. Please nod or say 'yes' if you understand. If you disagree with any statement you may not continue with the experiment. You are allowed to end the experiment and leave at any time – without providing any reason or explanation."

<Waits for answer.>

"There are no risks associated with this experiment."

<Waits for answer.>

"There is the possibility to feel slight discomfort or fatigue from looking intensely at the computer screen."

<Waits for answer.>

"All results are anonymized, and no confidential data are kept."

<Waits for answer. If a participant refuses any of these rights, the participant cannot continue any further with the experiment.>

"Now that you are aware of your rights and have agreed to continue to the experiment phase, please read and sign the consent form.

<Hands consent form and waits for the participant to fill it out. Takes the first copy for us, gives the second copy to the participant.>

"Now please fill out the demographics form."

<Hands demographics form and waits for the participant to fill it out. Once completed, writes the gender on the top of the form.>

"I will now describe the procedure step-by-step."

8.8. Task Descriptions (for participants)

"You will begin with three practice estimations to help you get at ease with using the program. Once you begin, a random graph will appear, and you will need to adjust the slider until you think the trend line fits the data on the graph. Once done, press "Next", and the next graph will appear. If you click on the button "Next" by mistake, you can return to the previous task by clicking on the button "Previous". Please perform the trial and let me know once you have finished. Now you may click on the "Start the practice" button."

<Waits until participant has completed the trial run.>

"Do you have any question regarding the use of the software?"

<Answers the participant's questions, if any.>

"The main experiment will be the same, as the practice, just with different parameters. You will have 96 normal tasks and 4 validation tasks in a random order. These validations tasks show the correct answer on the screen, and you just need to move the slider to fit it accordingly. You are now ready to begin the main experiment. Please call me if you have any questions or problems during the experiment. You may press “Start the main experiment” once you are ready. Good luck!”

8.9. Final Debriefing

“The experiment is now over. Please tell me how the experiment went for you.”

<Writes answers.>

“If you have any additional questions or comments you would like to share regarding the experiment or topic at hand, feel free to do so now.”

<Answers participant’s questions, if any.>

“Thank you for participating. Here is a bar of chocolate to compensate for your time and effort.”

<Gives chocolate to the participant.>

“Have a nice day!”

8.10. Data Collection

We do not record any audio, video or taking any pictures. We will have the following data collections:

- Demographic questionnaire;
- Software data output;
- Notes taken during the experiment.

9. APPENDIX

9.1. Appendix 1 – Consent From

Informed consent form

Participant ID: ____________________

Title of the study: Replication of "Regression by Eye: Estimating Trends in Bivariate Visualizations" by Michael Correll and Jeffrey Heer

Investigators:
Bianca Stancu | bianca.stancu@uzh.ch | University of Zürich
Alix d'Agostino | alix.dagostino@uzh.ch | University of Zürich
Natalia Obukhova | natalia.obukhova@uzh.ch | University of Zürich
Tim Brlan | tim.brlan@uzh.ch | University of Zürich
Alexander Pfyffer | alexander.pfyffer@uzh.ch | University of Zürich
Mathias Lüthi | mathias.luethi@uzh.ch | University of Zürich

Introduction.
Thank you for considering taking part in the research study on regression by eye. We ask you to carefully read this form and ask all the questions that may arise before agreeing to participate in the study. You were selected as a possible participant because:

1. You are over 18 years old;
2. You use computers regularly;
3. You have not read "Regression by Eye: Estimating Trends in Bivariate Visualizations" by Michael Correll and Jeffrey Heer
4. You do not have visual impairments which can affect your ability to see graphs on the computer screen

If any of these statements above are false, please inform an investigator immediately.

Purpose of the study.
The main research question of the study is how different features of the graph affect the estimated trend of the bivariate visualization by an individual. The research and its results can be published and used for the research purposes in the University of Zürich.

Description of the study procedures.
If you qualify and accept to partake in this study, you will be given a brief introduction to the topic, complete a demographics questionnaire, read the instructions for the tasks in this study, complete four trial estimations and then continue with the main 96 trials plus 4 validation tasks. All the tasks will be completed on a computer, where each trial will take approximately 5 to 10 seconds. After completing the study, you will be able to ask any questions regarding the tasks performed or the topic and given a compensation in the form of a chocolate bar. Your participation will take between 10 and 20 minutes depending on how quickly you answer.

Risks, discomforts and rights.
Please, carefully read the information below. Make sure you understand your rights fully and feel free to ask any questions.

- The participants are allowed to end the experiment and leave at any time – without reason or explanation.
- There are no risks associated with this experiment.
- There is the possibility to feel slight discomfort or fatigue from looking intensely at the computer screen. You are asked to immediately inform an investigator if you do.
- All results are anonymized, and no confidential data are kept.
- You will work with the computer, analyzing graphs.
- All the data collected will be purged after 5 years from the collection date.

- You can contact the investigators mentioned above at any time and withdraw your data from the study even after its completion – without any justification.

Consent

1. I, ______________________________________, confirm that I have read, understood and agree to all the information above and all my questions have been answered.
2. The compensation has been discussed with me and I understand that I cannot ask for any other compensations now or in the future.
3. I understand that my participation is fully voluntary, and I can leave at any time without giving reason or receiving any kind of punishment.
4. I understand that my confidential information will not be stored, and all the data will be anonymized.
5. I confirm that I am 18 years old or older and not a part of the vulnerable population.

First and last name of the participant: _____________________________ Signature: _________________

First and last name of the investigator: ____________________________ Signature: ________________

Date: _____________________

9.2. Appendix 2 – Demographics Questionnaire

Demographics questionnaire

Participant ID: ________________________

Title of the study: Replication of "Regression by Eye: Estimating Trends in Bivariate Visualizations" by Michael Correll and Jeffrey Heer

1. How old are you? ____

2. Which gender are you?

- ☐ Female
- ☐ Male
- ☐ Other: ____
- ☐ Prefer not to answer

3. What nationality are you? ____

4. Do you have a degree or are you currently studying at university?

- ☐ Yes
- ☐ No

5. If you answered yes to the question 4:

What degree level are you pursuing, or have you received (e.g. BA, MA, etc.)?

____

In which faculty is or was your major?

____

6. Do you have a visual impairment which would prohibit you from seeing detailed graphs?

- ☐ Yes
- ☐ No

7. Do you use computers more than once a week?

- ☐ Yes
- ☐ No

8. Do you analyze graphs and/or tables more than once a month?

- ☐ Yes
- ☐ No

9. Have you read the paper "Regression by Eye: Estimating Trends in Bivariate Visualizations" by Michael Correll & Jeffrey Heer?

- ☐ Yes
- ☐ No

Date: ________________