

# Lab 1 - Redwood Data, Stat 215A, Fall 2019

Your name

September 04, 2020

Please use this structure for your report, but you do not have to slavishly follow this template. All bullet points are merely suggestions and potential points to discuss in your writeup. Your report should be no more than 12 pages, including figures. Do not include *any* code or code output in your report. Indicate your informal collaborators on the assignment, if you had any.

## 1 Introduction

Things to potentially include in your introduction:

- Describe the premise of your exploratory data analysis and put your analysis in the domain context
- Explain why studying this redwood data interesting and/or important
- What are the implications of better understanding this data?
- What is the purpose of your exploratory data analysis?
- Outline what you will be doing in the rest of the report/analysis

## 2 Data

- What is the data that you will be looking at?
- Provide a brief overview of the data
- How is this data relevant to the problem of interest? In other words, make the link between the data and the domain problem

### 2.1 Data Collection

- How was the data generated?
- Discuss the measurements of each variable in the data that you will be using in your analysis

### 2.2 Data Cleaning

- Discuss all inconsistencies, problems, oddities in the data (e.g., missing data, errors in data, outliers, etc)
- What steps did you take to clean the data, and why did you clean the data in that way?
- Record your preprocessing steps in a way such that if someone else were to reproduce your analysis, they could easily replicate and understand your preprocessing
- You may find it helpful to include relevant plots that help to explain the choices you made when cleaning the data
- Be transparent! This allows for others to read your work and make their own educated decisions on how best to preprocess the data.

### 2.3 Data Exploration

- The main goal of this section is to give the reader a feel for what the data “looks like” at a basic level

- Think about plots that summarize the data, plots that convey some smaller findings which ultimately motivate your main findings
- A good report will tie everything together so that there is a reason for every figure in the story

## 2.4 Reality Check

- Do a reality check. What reality could you compare your cleaned data to?
- Clearly state your assumptions and explain why this reality check is useful.
- Does your cleaned data pass the reality check or are there issues? Discuss.

## 3 Graphical Critique

- Critique the plots in Figures 3 & 4 in Tolle et al.
- What questions did they try to answer? Did they answer them successfully?
- Did they raise any questions not addressed in the text?
- Would you change them at all?

## 4 Findings

- Present three interesting findings and produce a publication quality graphic for each along with a short caption of what each shows.
- Don't forget to appropriately label axes, titles
- Think carefully about use of color, labeling, shading, transparency, etc.
- Also interpret and provide an insightful discussion of what your figures show

### 4.1 First finding

Describe it and place a figure here

### 4.2 Second finding

Describe it and place a figure here

### 4.3 Third finding

Describe it and place a figure here

### 4.4 Stability Check

Take one of your findings and present a perturbed version. How does this affect your finding? Add a before and after plot here.

## 5 Discussion

- Did the data size restrict you in any way? Discuss some challenges that you faced as a result of the data size.
- Address the three realms: data / reality, algorithms / models, and future data / reality.
- Where do the parts of the lab fit into those three realms?
- Do you think there is a one-to-one correspondence of the data and reality?
- What about reality and data visualization?

## **6 Conclusion**

- You should make attempts to connect your findings/analysis back to the domain problem in every section of this report, but here in the conclusion, you can reiterate your main points and provide overarching remarks on the redwood data as it relates to the domain problem

## **7 Academic honesty statement**

Please address to Bin.

## **8 Bibliography**