# Homework3 @DA2210

Rakin Ali

September 28, 2023

## Contents

# 1 Two research studies

This section covers two research studies written in the assignment description.

## 1.1 Does comment improve code?

**Can we conclude that comments improve code? Give suggestions to improve the study:**
Strong arguments can be made for a yes but also a no. My take on this question is that a more nuanced analysis is needed.
The study, in my opinion, shows a correlation between commenting code and producing better code however does not really show causation. Factors such as the coding skills of the students or the difficulty of the problem they were facing could have had an impact.
Since group A were following a style guide, they might have written code which produced better results than if they were not following the style-guide. In other words, not commenting code might have produced better results than what it should have?
The study does not mention how many students, their educational background, their coding background, their familiarity of the problems and other stuff. This means that the sample size can also affect the result of the studies.
The term "better code" needs to be defined more precisely so that it preferably can be quantified and evaluated better. Does better code mean readable code or code that is error-free but hard to read?
I could go on and on for a long time. To summarize, this study may strengthen previous literature that comments improve code but this study on it's own cannot be used to conclude anything. If someone made this conclusion, I'd have more questions than answers so to say.

## 1.2 Haskell versus Java

This was slightly more difficult as Java is actually quite fast but obviously slower than C++ programming language. That being said, the exact same arguments and most of the improvements for 1.1 can be made for experiment as well. This study on it's own cannot be used to conclude anything, just strengthen previous literature or go against it.

# 2 Zips's law

This section discusses Zip's law and applies it to a dataset.

## 2.1 Dataset chosen

The dataset chosen was a word frequency list as it was easier to work with and I simply wanted to complete this assignment fast. The code can be viewed here
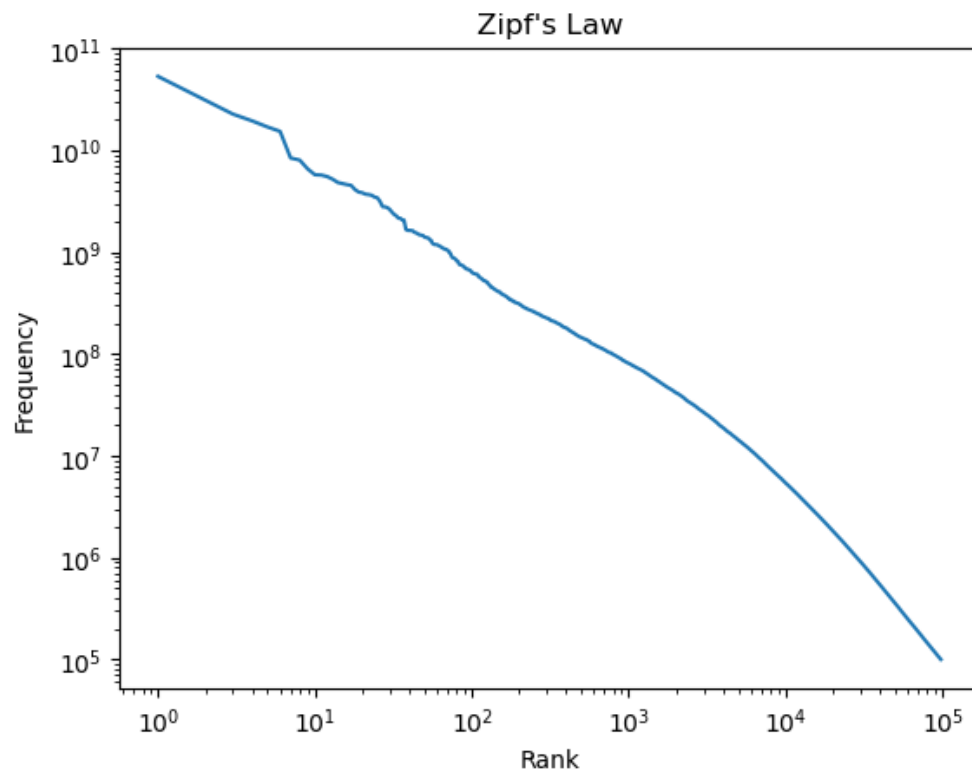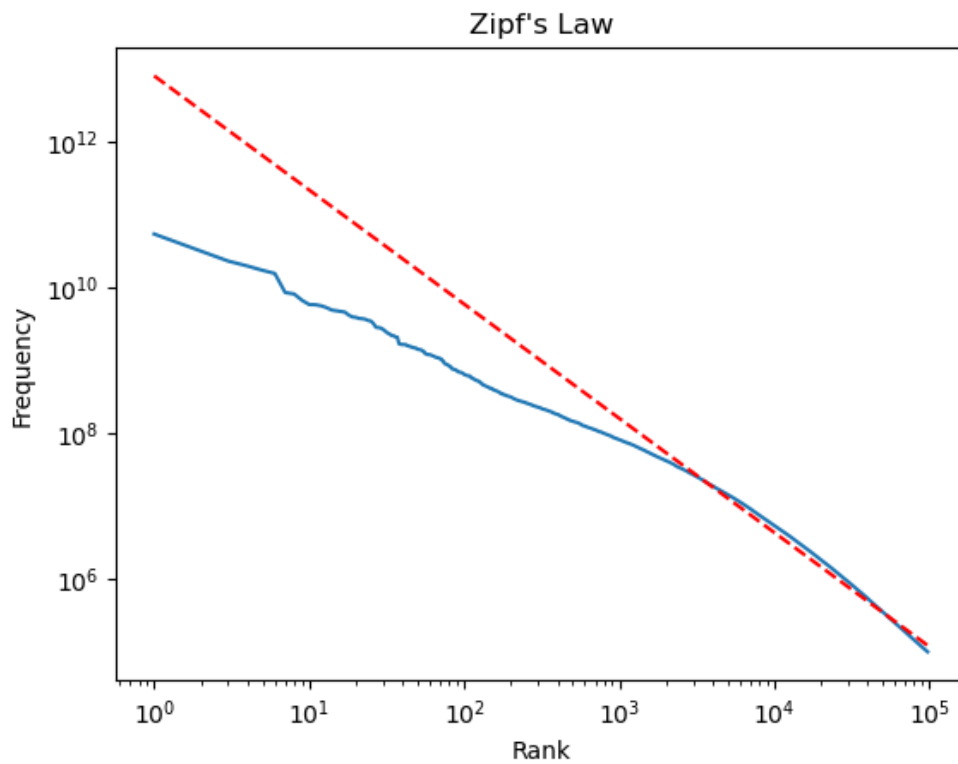
Figure 1: A plot on word frequency list and their rank.

## 2.2 B and C: ORder the dataset and use linear regression

Figure 2: A log plot on word frequency list and their rank.