Assignment 2
Interpreting Graphics

Applied Data Science

Instructions

This assignment has three (3) parts.

Submit your assignment as a .pdf file with your name and email in the header, single-spaced and font no-larger than 12pt. There is no need to include the graphs in your assignment. Assignment will be graded on the quality of the descriptions and overall quality.

Part 1

You will be presented with four (4) graphs from scholarly journals in the fields of data science and informatics. For each, use the skills you have learned in the lectures and your previous experience to describe the relevant message each graph conveys. As a reminder, the textual description should cover all the most important elements of the graph. Your description for each graph should not be more than two paragraphs.

Part 2

You should write two (2) paragraphs: one (1) paragraph explaining which graph you thought was the most communicative (i.e., which graph communicated its information the best); and one (1) paragraph describing which graph you thought was the least communicative or most-erroneous.

Part 3

You will perform an exploratory data analysis of the low birth rate data set. You may wish to modify the R script provided from the walkthrough, use R Commander (Rcmdr), or import the data into Tableau for this assignment. Consider which variables in this data set are most important. How are those variables distributed? Provide at least 2 graphs that helped you get a feel for the data and a one paragraph discussion of your findings.
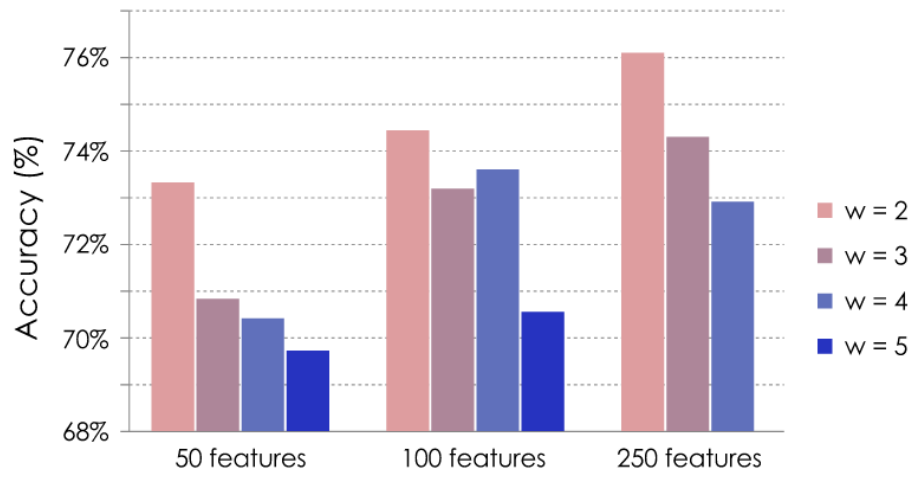
Graph 1

Figure 1: Aidos, Fred, Initiative, et al.
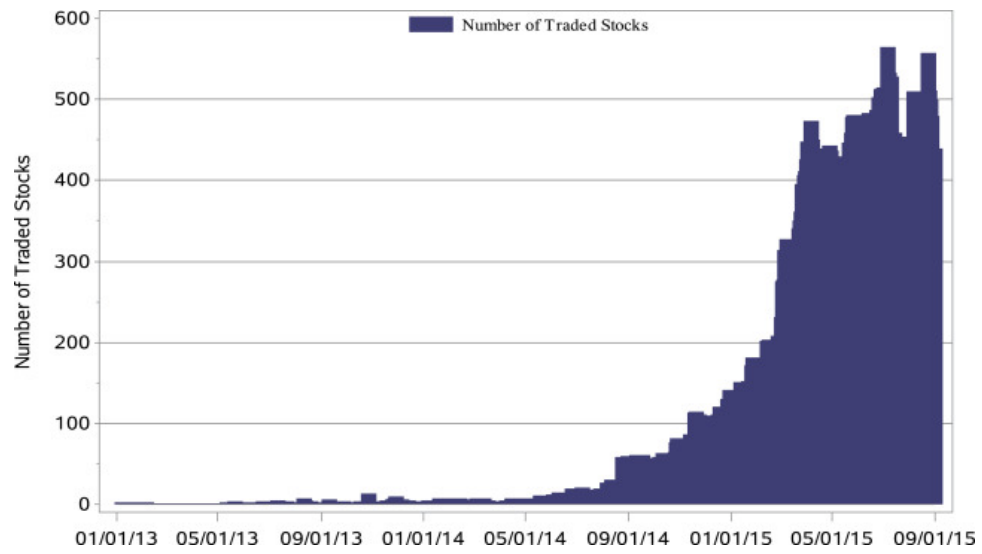
Graph 2

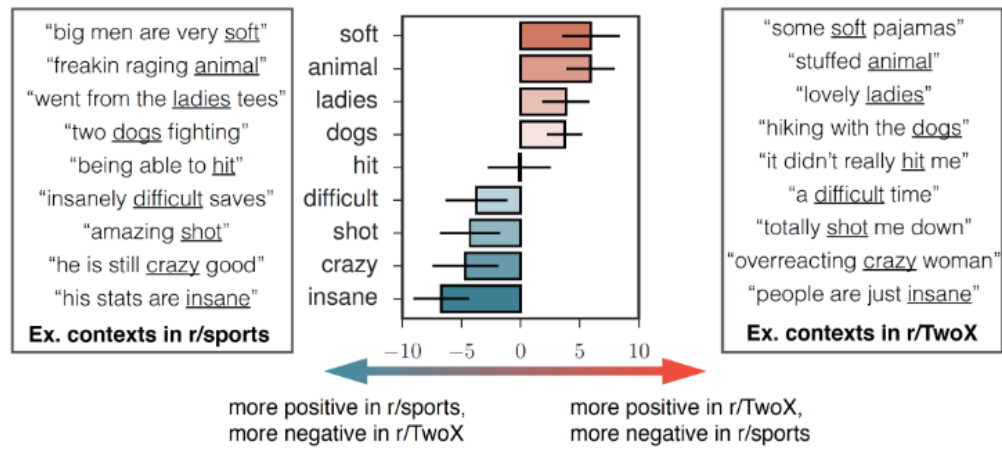Figure 2: Li and Qiao

Graph 3

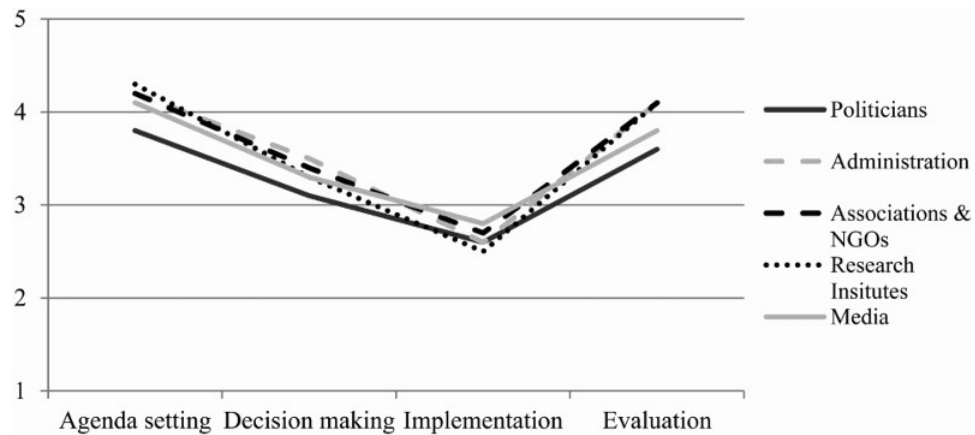Figure 3: Hamilton, Clark, Leskovec, and Jurafsky

Graph 4

Figure 4: Fawzi

3

References

[1] Helena Aidos, Ana Fred, Alzheimer's Disease Neuroimaging Initiative, et al. "Discrimination of Alzheimer's Disease using longitudinal information". In: *Data Mining and Knowledge Discovery* (2017), pp. 1–25.

[2] Nayla Fawzi. "Beyond policy agenda-setting: political actors' and journalists' perceptions of news media influence across all stages of the political process". In: *Information, Communication & Society* (2017), pp. 1–17.

[3] William L Hamilton et al. "Inducing domain-specific sentiment lexicons from unlabeled corpora". In: *arXiv preprint arXiv:1606.02820* (2016).

[4] Weiping Li and Gaoxiu Qiao. "Volatility and returns of the New Third Board market in China". In: *The Journal of Finance and Data Science* (2017).