DMJN328 Final Exam

Simon J. Kiss

April 17, 2020

1. What is meant by the data-to-ink ratio in data visualization (Healy 2018)? Why is it a good thing? What are some practical reasons for why it might occasionally have to be abandoned?

2. What is meant by a tidy data-set? How many variables are in this graph:

3. Imagine the following data set below. How would you reshape the data so that it is tidy? Demonstrate by drawing in the space below, copying-and-pasting and reshaping in Microsoft Word.

Province	2002	2003	2004
Toronto	4	5	6
Brantford	4	5	6
Hamilton	3	5	6
Waterloo	7	5	6

5. What is meant by levels of measurement? What are the broad categories of levels of measurement in statistical analysis? Why do they matter for visualization?

6. What are three measures of central tendency? What are the differences between them? What are some strengths and weaknesses of each?

7. What is meant by the normal distribution?

8. In words, what is the general relationship between sample size and standard error?

9. What is a 95% confidence interval (or margin of error)?d

10. What does a p-value communicate? What are some disadvantages of p-values? What does a p-value *not* communicate?

11. What is the difference between a cross-sectional study and an experimental study? What advantages and disadvantages do both offer?

12. What is meant by publication bias in scientific publishing?

References

Healy, Kieran. 2018. *Data Visualization: A Practical Introduction.* Princeton University Press.