Final Exam

Simon J. Kiss

01/04/2020

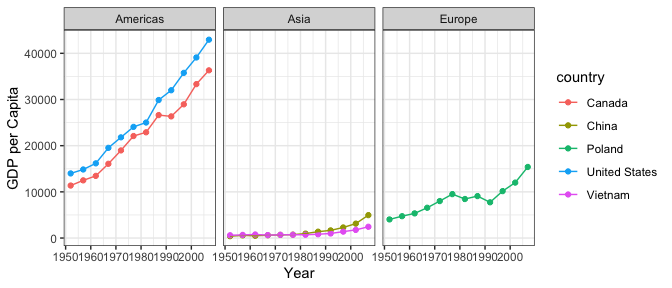
# Instructions

Answer *nine* of the following 12 questions in paragraph-length. Draw on any material you want (e.g. the course texts, lecture materials, course slides). Each question will be graded out of five, on the following scale.

|  |  |
| --- | --- |
| Points | Definition |
| 5 | Outstanding and comprehensive |
| 4 | Very good and nearly comprehensive |
| 3 | Sufficient and shows basic understanding |
| 2 | Not sufficient but the answer demonstrates some understanding of material close to the answer |
| 1 | There is a meaningful answer that has at best a tangential relation to the material |
| 0 | false with no demonstration of understanding the material |

# Questions

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. How many variables are in this graph:



1. What is meant by a tidy data-set? Imagine the following data set below that shows the pretend unemployment rate in three Ontario cities. How would you reshape the data so that it is tidy? Demonstrate by making a new, tidier table next to it or below it in your word processor.

|  |  |  |
| --- | --- | --- |
| City | 2002 | 2003 |
| Toronto | 4 | 5 |
| Brantford | 4 | 5 |
| Hamilton | 3 | 5 |

1. 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?
2. What are three measures of central tendency? What are the differences between them? What are some strengths and weaknesses of each?
3. What is the central limit theorem? What does it enable?
4. In words, what is the general relationship between sample size and standard error? Is a larger sample size always better?
5. What is a 95% confidence interval (or margin of error)? What happens to a confidence interval (or margin of error) as the sample size increases? What is the logic behind using a 95% confidence interval, as opposed to a 90% or 80% confidence interval?
6. What does a p-value communicate? What are some disadvantages of p-values? What does a p-value *not* communicate?
7. What is the difference between a cross-sectional study and an experimental study? What advantages and disadvantages do both offer?
8. What is meant by publication bias in scientific publishing?

# References

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