



Module 9 Practice Problems: Introduction to Statistics and Data Visualization

1. Identify the *Description*, the *Unit of Measurement* and the *Unit of Observation* for the following variables.

- (a) The weekly average price of a Bitcoin.
- (b) The number of babies born in New Amsterdam hospital daily.
- (c) The average number of minutes a Netflix subscriber watches Netflix in a week.
- (d) The daily precipitation level in New York.
- (e) The daily Canada/US exchange rate.

2. Determine if the following variables are qualitative (categorical) or quantitative (numeric). If categorical, are they nominal or ordinal? If numeric, are they discrete or continuous?

- (a) The number of snowy days in Toronto in 2022.
- (b) The area code of a respondent's telephone number.
- (c) The weight of a newborn.
- (d) The time a customer spends in Costco.
- (e) Responses to the question "How would you rate the food in this restaurant?", with choices from 1 star (*horrible*) to 5 star (*excellent*).
- (f) The make of cellphone owned by a student.
- (g) Fraction of the US population that is foreign-born.
- (h) The letter grade a student receives in a course.

Data for the following questions is contained in the Excel (or .ods) file GAMES_StatIntro_PracticeQuestion.xls that can be found in the Module. If needed, the file can be exported to Google Sheets.

3. The Excel sheet "Olympics medals" contains the medals won by the top 10 countries in the 2020 Tokyo Summer Olympics.

- (a) Create a bar or column chart for the total number of medals won by each country. Do you prefer a horizontal bar or a vertical one?
- (b) Create a bar or column chart separately for the gold, silver and bronze medals for each country.
- (c) The sheet also contains the 2020 population for each country. How would one compare the per-capita medal performance for each country? How do the countries compare according to this measure?

4. The Excel sheet "Cell phones" contains the results of a survey of 40 individuals about the make of their cell-phone.

- (a) Create a frequency distribution for this data.
- (b) Create a bar chart for this data.
- (c) Create a pie-chart for this data to show the percentages of the various brands. Make sure that the most popular option "stands out". *You will often see such charts in the media to show market shares of different companies.*

5. The Excel sheet "Student survey" contains the results of a survey of 60 students about a course they took.

- (a) What is the type of the variable here?
- (b) Create a frequency distribution for this data.
- (c) Create a bar or column chart for this data. Do you prefer a horizontal bar or a vertical one?
- (d) Create a pie-chart for this data.

6. Reconsider the "Olympics medals" data. Now, instead of looking at the total number of medals by country, let us look at the number of gold, silver and bronze medals won by each country. How would one present this data as a grouped bar/column chart? In drawing this chart, take into consideration what is the question that you are trying to shed light on via this chart.

7. The Excel sheet "Cdn elections 2021" contains data from the 2021 Canadian Federal elections for the four largest provinces: Ontario, Quebec, British Columbia and Alberta. It gives data on the vote shares of the major parties and also the number of seats they won in these provinces.

- (a) Create a grouped bar or column chart for the data on vote shares. Which grouping variable are you choosing? Why?

- (b) Suppose instead you created separate pie-charts for each province to depict the distribution of votes among the parties. In terms of visualization, how do these compare with the grouped column/bar chart in (a)?
- (c) Create a grouped bar or column chart for the data on seats won.

8. T.A. Boden, G. Marland and R.J. Andres have estimated global, regional and national fossil-fuel CO₂ emissions over the period 1950-2014. In the Excel sheet "Carbon Emissions", you will find their year-by-year data on per-capita CO₂ emissions for the entire world, for North America, Western Europe and Africa.

- (a) Plot the time trends in per-capita CO₂ emissions separately for the entire world and for each of the three regions. What do you notice from these plots?
- (b) Then plot the time trends for the three regions in the same plot. What do you notice from the plot?

Reference: Boden, T. A., Andres, R. J., and Marland, G. Global, Regional, and National Fossil-Fuel CO₂ Emissions (1751 - 2014) (V. 2017). United States: N. p., 2017.
Web. doi:10.3334/CDIAC/00001_V2017.

9. The Dow Jones Industrial Average (DJIA) and the NASDAQ are two widely followed US stock market indices, while the Toronto Stock Exchange (TSX) is the most prominent Canadian stock market index. In the Excel Sheet "Stock Market Indices", you will find the closing value (i.e. the end of the day value) of these indices over the period December 1, 2021 to January 31, 2022.

- (a) Plot separately the time trends for the three indices over this period.
- (b) What do you see about their behavior over this period? Did all three indices behave similarly?

10. Bitcoin and Ethereum are two of the most prominent cryptocurrencies in the market these days. In the Excel Sheet "Cryptocurrencies", you will find the closing value (i.e. the end of the day value) of Bitcoin and Ethereum (in US Dollars) over the period July 1, 2021 to December 31, 2021.

- (a) Plot the time trends for the price of the two cryptocurrencies. What do you see about their behavior over this period? Did both behave similarly?
- (b) Draw a scatter plot of Ethereum versus Bitcoin closing prices. Be sure to label the axes and adjust the axes to remove unnecessary white space from the graph. Do you notice any relationship between the two?

11. Let us now investigate if there is any relationship between the DJIA, the NASDAQ and the TSX.

(a) Draw a scatter plot of the DJIA versus the NASDAQ. Be sure to label the axes and adjust the axes to remove unnecessary white space from the graph. Do you notice any relationship between the two?

(b) Draw a scatter plot of the DJIA versus the TSX. Do you notice any relationship between the two? Is this similar to that between the DJIA and the NASDAQ?

(c) Draw a scatter plot of the NASDAQ versus the TSX. Is this relationship similar to that between the DJIA and the TSX?

12. The Excel sheet "Course Marks" contains the marks of students in the midterm and the final exams of a course.

(a) Draw a scatter plot of the midterm marks versus the final marks. Do you notice any relationship? In particular, do students who do better on the midterm also tend to perform well on the final exam?

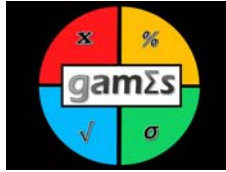
(b) Generate a histogram of the midterm marks using Excel's automatic function. Try to adjust the bins according to your criteria. In particular, note that anyone scoring less than 50% on an exam is deemed to have failed.

(c) Generate a similar histogram for the final exam marks. Is the distribution of the final marks similar to that for the midterm?

13. The Excel sheet "US Crime Rates" contains the violent crime rate per 1 million residents for each of the 50 US states and the District of Columbia for 2018 and 2011.

(a) Generate a histogram of the 2018 crime rates, using Excel's automatic function. Try to adjust the bins according to your criteria.

(b) Generate a similar histogram for the 2011 crime rates. Is the distribution of the crime rates similar across the two years?



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) <https://creativecommons.org/licenses/by-nc-nd/4.0/> Unless otherwise noted, all content in this video was created by Catherine Pfaff, Sumon Majumdar, and Robert J. McKeown. You are free to copy and share this material in any format, but you must give appropriate credit to the authors. This project is made possible with funding by the Government of Ontario and through eCampusOntario's support of the Virtual Learning Strategy. To learn more about the Virtual Learning Strategy visit: <https://vls.ecampusontario.ca/>.