Assessment 1

Note: You can use whatever software you prefer to do these questions. If Inzight can be used, use Inzight, else use whatever works.

1. Let's figure out which state in the United States do I want to go to if I want to minimize my risk tolerance for being murdered by a gun. Find the data here: https://en.wikipedia.org/wiki/Gun_death_and_violence_in_the_United_States_by_state . You will need to turn it into a CSV file and upload it to Inzight (or Excel).
    a. Make histograms of each of the variables. Calculate the mean and median of the gun homicide rate for all the states. Which state has the minimum, while still having some data? (using identify points will make Inzight crash – instead just identify the minimum using the summary and find it in your dataset, or do something else to figure this out).
    b. What is the odds ratio between homicide rate in the USA vs the Netherlands? You will need to look up and cite where you find the gun homicide rate for both. Interpret this value. Could you have used the mean from all the states? Treat homicide rate as an expected frequency.
    c. Can you generate a table of the odds for the best 5 states where you have values for gun homicide rate? Treat gun homicide rate as an expected frequency.

2. Download the uploaded dataset on Canvas for Gapminder. You will need to filter the data on country and year to get the plots below and the mean and standard deviation.
    a. Generate the mean and standard deviation for the Netherlands for life expectancy and infant mortality from 2000 to 2015.
    b. Make a line plot of the infant mortality between 2000 and 2015. Use the join points functionality in Trend Lines and Curves.

3. Make some good visualizations.
    a. Inzight Lite (or elsewhere): Make a plot of age decade against BMI for the NHANES dataset (nhanes_2000 in Future Learn – can download it there if you want). Make the dots be filled and slightly transparent alongside the box plot. In a separate figure, create a violin plot, no need for dots.
    b. Inzight Lite (or elsewhere): From the same NHANES dataset:
        i. Make a visualization that helps you see the distribution of household incomes (variable HHIncome) across races (variable Race1).
        ii. What percentage of White people and Mexican people have an income between 20k to 25k and what percentage have an income between 65k to 75k relative to the total number of White/Mexican households? Give exact values.
    c. Gapminder Tools: Make trend lines using the GapMinder Tools of the fertility rate over time (so you need to change the x-axis from the default) for your

choice of country. Reduce opacity in Options for the plot for the non-selected countries. You can screenshot to get the image or download as SVG if you're able to display that in a document.

   d. Gapminder Tools: Make a plot for the infant mortality rate for two countries that defies your expectation. Write down your expectation and justify it in a line or two.
   e. Gapminder Tools: Find 2 countries that show unexpectedly high happiness scores (look under the tab "Society" in GapMinder) for their GDP per capita? Mark these and show them in a plot.

4. Make 3 bad visualizations using the tools we've discussed so far (Inzight Lite/Gapminder Tools) or any other tools (Excel, etc.). Or find some bad visualizations from the news in the past year. Justify each inclusion. Go through https://viz.wtf/ and Lecture 2. This shouldn't only be 3D pie charts.

5. What is the data, sample, study population and target population for the following cases:
   a. How is gun violence in the USA changing over time?
   b. How can I judge if the birth length of a baby is abnormal?

6. Go over the NHANES dataset or any other dataset on Inzight and produce two nonsense correlations (they need to be non-zero) between a pair of numeric variables and associated scatterplots and trend lines. Explain what could be driving those correlations, that is, what is the potential lurking factor(s), if any in 1 or 2 lines?

7. Design a randomized control trial to assess whether large language models are making everyone in the university a little bit more incapable of thinking for themselves (critical reasoning). Explain how you would set up the sampling, what the treatment/control would be and how you could assess the change in critical reasoning across groups.

8. Go find a recent news report (within the past month) that includes a claim based on statistical evidence and critically assess the claim: What is the claim? What is the evidence provided to support the claim? Have they overstated the claim? (here's an example: https://www.bbc.com/news/articles/cj90m4errlgo , search for cancer/other diseases to find more articles like this or go to the Health page of news websites). Check out the textbook/slides to understand how to critique claims.