Advanced Visualization

Use this LBDE to gain experience with advanced visualization techniques in R / Python.
ULTIMATE GOAL: try to gain a better understanding of Canadian health regions.

Dataset

HR_2016_Census_simple.xlsx

Problem Description

The population of Canada is divided physically into provincial and territorial areas, most of which are further subdivided into **health regions**.

Census information (from 2016) is available for those health regions

- https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710012201
- https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710012301

The equivalent 2018 dataset has been clustered to produce **peer groups**: the result is shown at

- https://www150.statcan.gc.ca/n1/pub/82-402-x/2018001/maps-cartes/rm-cr14-eng.htm

In this lab, you will explore and visualize the 2016 dataset in various ways using ggplot2 and dashboard tools (the goal is not to reproduce the map).

1. Load the data and **summarize** it (extract the rows with a 4-digit geocode). Consult the online files to help your data understanding.

2. Clean the data and **impute** missing values (if necessary).

3. Create 10+ *ggplot2* **univariate visualizations** for the dataset: 5+ using your understanding of the context to select features, and 5+ by selecting features randomly.

4. Create 10+ ggplot2 **multivariate visualizations** for the dataset: 5+ using your understanding of the context to select features, and 5+ by selecting features randomly.

5. Using your findings in steps 3. and 4., design a **dashboard** to inform stakeholders of the situation in the various health regions. Remember to ask the important questions about the dashboard. You may implement your design in Power BI or Tableau, should you wish to do so.