

# lab2

## # Lab Exercises

To be handed in via submission of quarto file (and rendered pdf) to GitHub.

1. Using the ‘delay\_\_2022’ data, plot the five stations with the highest mean delays. Facet the graph by ‘line’
2. Using the ‘opendatatoronto’ package, download the data on mayoral campaign contributions for 2014. Hints:
  - find the ID code you need for the package you need by searching for ‘campaign’ in the ‘all\_data’ tibble above
  - you will then need to ‘list\_package\_resources’ to get ID for the data file
  - note: the 2014 file you will get from ‘get\_resource’ has a bunch of different campaign contributions, so just keep the data that relates to the Mayor election
3. Clean up the data format (fixing the parsing issue and standardizing the column names using ‘janitor’)
4. Summarize the variables in the dataset. Are there missing values, and if so, should we be worried about them? Is every variable in the format it should be? If not, create new variable(s) that are in the right format.
5. Visually explore the distribution of values of the contributions. What contributions are notable outliers? Do they share a similar characteristic(s)? It may be useful to plot the distribution of contributions without these outliers to get a better sense of the majority of the data.
6. List the top five candidates in each of these categories:
  - total contributions
  - mean contribution
  - number of contributions

7. Repeat 5 but without contributions from the candidates themselves.
8. How many contributors gave money to more than one candidate?