

CSC 497 FALL/WINTER SESSION PROJECT PROPOSAL

Topic: The COVID-19 period: What Changed?

The discovery and widespread of COVID-19 led to drastic changes to many Canadian lives: students especially.

Current paradigm:

Some students experienced reduced hours of work at their hourly-wage jobs in addition to the change from in-person to virtual access to education. Time series data and reports are continuously being produced and published by government agencies like Statistics Canada and the Public Health Agency of Canada, and these data can be explored to report on time series trends or changes within the period of Mar. 2020 to Mar. 2021.

Gap:

Lack of comprehensive news and reports on changes that have occurred during the period of Mar. 2020 to Mar. 2021 due to the impossible challenge of answering every question on the impact the COVID-19 Pandemic had on Canadians.

Research:

To conduct a research on “what changed” during the COVID-19 period, we will be exploring how the COVID-19 pandemic has impacted the working hours of hourly paid employees in Canada. The research will be conducted by retrieving data from Statistics Canada and exploring trends in hourly paid employees working hours before and during the COVID-19 pandemic period.

Forecast:

To answer the research question will require producing research report deliverables, which will be released following a phased approach that allows for an in-depth research, data analysis and review to be conducted on the selected research question. The phases are namely:

Phase 1 [Week 1 of the Semester]: Data preparation, cleaning, research and analysis

[Week 1-2]: Data preparation and cleaning

[Week 3-4]: Research and analysis

Phase 2 [Week 5 of the Semester]: Report rough draft

Phase 3 [Week 8 of the Semester]: Conference poster

Phase 4 [Week 9 of the Semester]: Poster presentation

Phase 5 [Week 12 of the Semester]: Structured abstract and Final Report

Proposition:

I will explore the impact of the COVID-19 pandemic on average weekly hours for employees paid by the hour. I expect to accomplish this by employing the academic research, data analysis and presentation skills I have acquired during my studies at the University of Victoria (UVic). The experiment will be performed by carrying out statistical data analysis on the data sets using a non-parametric two-sample statistical test for equality and Z test for equality of means (Appendix A). The results of the experiment will be compiled into a final report providing recommendations and next steps for the project. I will use a conference poster and a research report to inform the UVic (student) community on relevant changes that has occurred during the pandemic and propose next steps.

Data and data questions

The data is available publicly from [Average weekly hours for employees paid by the hour, by industry, monthly, unadjusted for seasonality - Open Government Portal \(canada.ca\)](#) and [Public Health Infobase - Data on COVID-19 in Canada - Open Government Portal](#) both as a .csv file. There are 1048576 data rows and 16 data dimensions in the first time series dataset published by Statistics Canada. The data dimensions are REF_DATE, GEO, DGUID, Overtime, North American Industry Classification System (NAICS), UOM, UOM_ID, SCALAR_FACTOR, SCALAR_ID, VECTOR, COORDINATE, VALUE, STATUS, SYMBOL, TERMINATED, and DECIMALS. The dimensions of interest here are REF_DATE and VALUE, which stand for reference date and average weekly hours for employees paid by the hour, respectively.

There are 7453 data rows and 17 data dimensions in the second time series dataset published by Public Health Agency of Canada. The data dimensions are PRUID, PRNAME, PRNAMEFR, DATE, NUMCONF, NUMPROB, NUMDEATHS, NUMTOTAL, NUMTESTED, NUMRECOVER, PERCENTRECOVER, RATETESTED, NUMTODAY, and PERCENTTODAY. The dimensions of interest here are DATE and NUMCONF, which stand for report date and number of confirmed cases, respectively.

Questions:

Some of the questions I hope to answer are, is the average weekly hours for employees paid by the hour for year 2020 different from the mean all-time average weekly hours for employees paid by the hour. Is the average weekly hours for employees paid by the hour for year 2020 greater or less than the mean all-time average weekly hours for employees paid by the hour provided by the dataset? Is the change in average weekly hours for employees paid by the hour an emerging pattern during pandemics? Are there warning signs that change to average weekly hours for employees paid by the hour is imminent? Is there a threshold at which further changes to the weekly hours for employees can be inferred?

Approach:

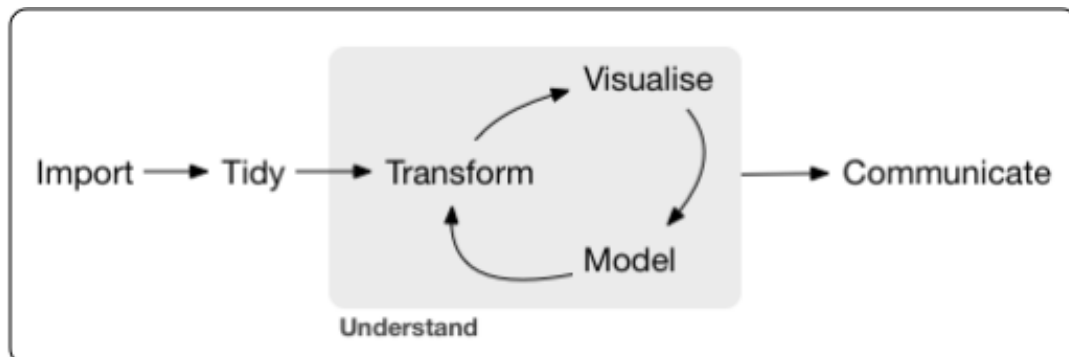


Figure 1. Stages of the data science project (Wickham & Grolemund, 2017)

Method

As shown in Figure 1 above, this data science project plan involves importing the data into R Studio and loading it into a data frame in R. Next, tidying the data to match the match the scope of the research and the collection and use of the data. Tidying the data will involve transforming the data into indicators we can visualize, analyze and use in answering our research questions through our research model. The analysis will require using custom built and R library functions for communicate stage of the data science project.

Risks:

Risk	Impact	Probability	Mitigation
Scope creep	Med	Med	Review of project scope and priorities with project supervisor
Inadequate allocation of dedicated human and tech. resources	Med	Low	Request that project supervisor allows for the acquisition and reallocation of resources needed
Project stakeholders cohesion	High	Low	The project stakeholders need to accommodate dynamic schedules and needs of its members

Table 1. Project risk assessment table

Project timeline:



Figure 2. Tentative project timeline (cropped snapshot of the much larger project Gantt Chart)

Access Gantt Chart at: <https://1drv.ms/x/s!ApNa20bP1ubhvGmZH2CHyc2TwFgQ?e=50AYjc>

References:

Wickham, H., & Grolemund, G. (2017). *R for data science: import, tidy, transform, visualize and model data*. O'Reilly.

APPENDIX A

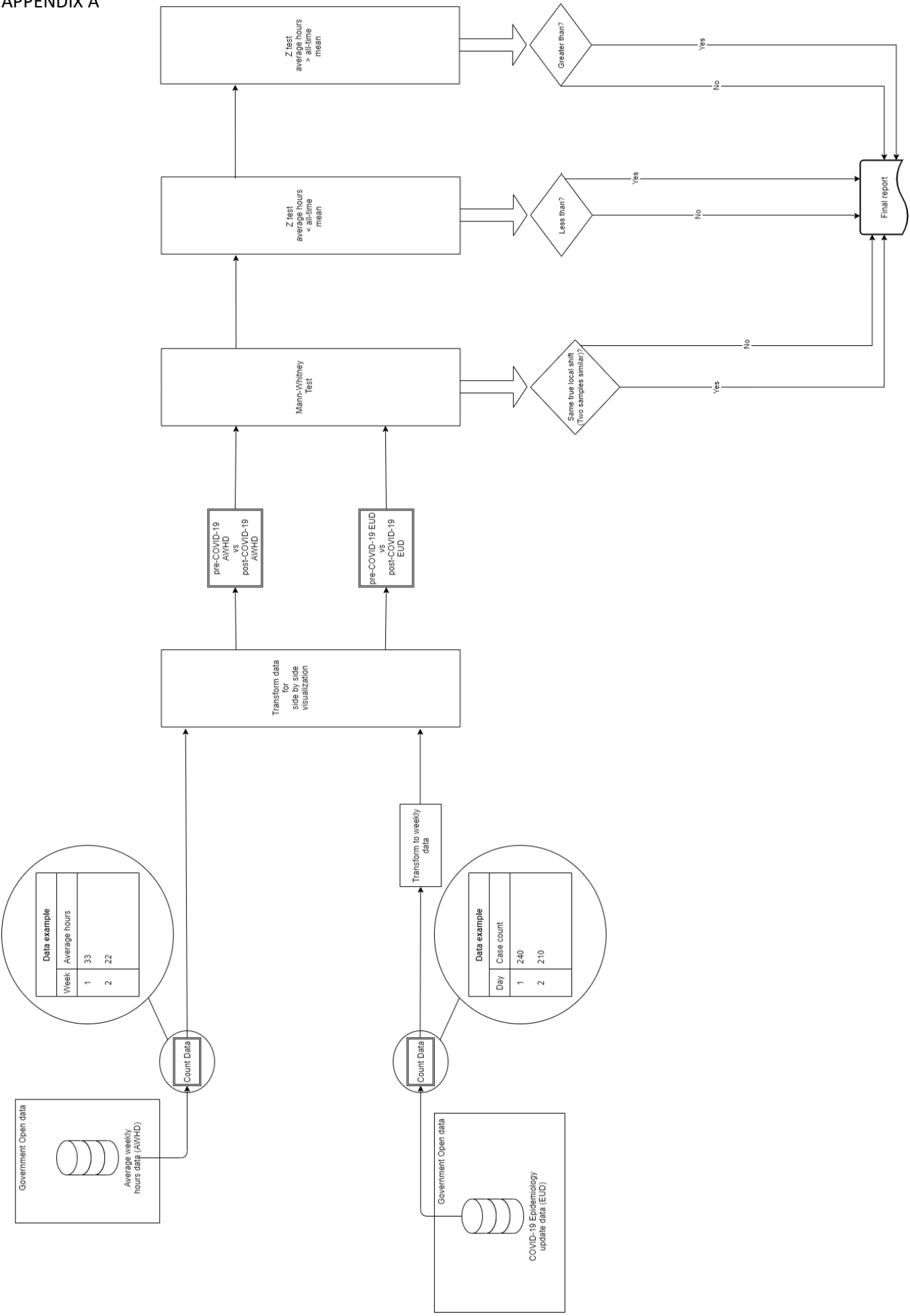


Figure 3. “The COVID-19 period: What Changed?” data science experiment plan