Module 3: R

Submodule 4: Manipulation

Expected length: 0.5 day

Guiding question: How can we manipulate a data table to display the most relevant information?

Concepts: filtering, selecting, arranging, mutating, the pipe, grouping, summarising, pulling, data cleaning

Description: This lesson introduces the student to the basic skills they will need to interact with tibbles and alter their data to prepare for further analysis or extract insights.

Instructor Preparation: Run all code to ensure it generates the same output as described in the lesson. Select dataset for data cleaning live coding.

| Materials and resources | Learning objectives |
| --- | --- |
| 04-manipulation\_deck.html | 1. Ability to filter, select, and arrange tibbles  2. Ability to create and change variables using mutate  3. Ability to extract insights from data using group\_by and summarise  4. Ability to apply the above techniques to clean data |

| Length | Lesson content | Guidelines, tips, and tricks |
| --- | --- | --- |
| 30min | Filtering (Wickham and Grolemund, 2017 Chapter 5, Timbers et al. 2021, Chapter 3.6)  Arranging (Wickham and Grolemund, 2017 Chapter 5)  Selecting (Wickham and Grolemund, 2017 Chapter 5, Timbers et al. 2021, Chapter 3.5) | Participants should be following along in their own RStudio environments. The concepts are introduced quickly so students can see them in use in the demos later. |
| 10min | The pipe (Wickham and Grolemund, 2017 Chapter 5 & 18; Timbers et al. 2021, Chapter 3.8) | Participants should be following along in their own RStudio environments. |
| 10min | Mutating (Wickham and Grolemund, 2017 Chapter 5, Timbers et al. 2021, Chapter 3.7, 3.10) | Participants should be following along in their own RStudio environments. The concepts are introduced quickly so students can see them in use in the demos later. |
| 30min | Summarising (Wickham and Grolemund, 2017 Chapter 5, Timbers et al. 2021, Chapter 3.9)  Grouping (Wickham and Grolemund, 2017 Chapter 5) | Participants should be following along in their own RStudio environments. The concepts are introduced quickly so students can see them in use in the demos later. |
| 60min | Cleaning (Alexander, 2022, Chapter 11) | Participants should be following along in their own RStudio environments. |
| 60min | Live coding: data cleaning | Ask students in advance for datasets of interest or use data that you are familiar with. Walk through cleaning the data live – ideally with little preparation beforehand, so students can see the process. Ask for contributions to make decisions during the process. |
| 60 min | Formative exercises  1. Filter the rows in the CES\_data dataset where the survey-taker is between 30 and 50 (cps19\_age).  2. Filter the rows in the CES\_data dataset where the survey-taker answered the cps19\_votechoice question (i.e. the cps19\_votechoice variable is not NA).  3. Select the variables cps19\_age and cps19\_province from the CES\_data dataset.  4. Select all variables except cps19\_province from the CES\_data dataset.  1. Create a variable in the dataset CES\_data that states if a person consumes news content or not (i.e. cps19\_news\_cons is equal to "0 minutes" or it is not).  2. Modify the variable cps19\_income\_number in the dataset CES\_data so that it is measured in thousands (i.e. divide the income number by 1000).  1. Use the CES\_data dataset. Group by cps19\_votechoice. Find both the median and mean rating of Trudeau (cps19\_lead\_rating\_23):  2. Use the CES\_data dataset. Group by cps19\_imm and cps19\_spend\_educ. Find the count for each group.  1. Fix this error:  ```{r, eval=F}  CES\_data %>%  summarise(mean = mean(cps19\_age)) %>%  group\_by(cps19\_gender)  ```  2. Fix this error:  ```{r, eval=F}  CES\_data %>%  filter(cps19\_vote\_choice == "Green Party")  ```  3. Fix this error:  ```{r, eval=F}  CES\_data %>%  mutate(cps19\_fed\_donate = factor(cps19\_fed\_donate,  levels = c("Yes",  "No",  "Don't know/ Prefer not to answer"))  ```  4. Fix this error:  ```{r, eval=F}  CES\_data %>%  select(cps19\_province  cps19\_age  cps19\_gender)  ``` | Allow 30 minutes of independent work time.  Take up for 30 minutes, with RStudio open to demonstrate.  Ask students for errors they encountered. |