

Lab Session 1: Introduction to R

Isabel Fulcher

3/2/2021

Set working directory

We will first set a working directory. Your working directory should be where you plan to save your R code and also where the example datasets have been stored. In the “Files” tab, navigate to this folder. Once you are there, select the “More” dropdown and select “Set As Working Directory”.

Install and load R packages

You should have already installed the tidyverse package. Now, you will need to load the package into R. This will allow you to use the functionality of the package.

```
library(tidyverse)
library(lubridate)
```

Load and view data

Load the outpatient visits “.rds” file and save it as a data frame called “outpatient”. An .rds file is an R object and is the best way to store data that will be used in R.

```
outpatient <- readRDS("session1_data/example_outpatient.rds")
```

View the first six observations in the outpatient datasets.

```
head(outpatient)
```

View the variable types in the dataset.

```
str(outpatient)
```

Load a .csv file in R and view the first six observations. Does anything look different about this data frame?

```
data_csv <- read_csv("session1_data/example_outpatient.csv")
head(data_csv)
```

Working with data

Filtering data

Filter the data to months that had greater than or equal to 8000 outpatient visits.

```
outpatient %>%
  filter(outpatient_visits >= 8000)
```

Filter the data to view outpatient visits in 2020.

```
outpatient %>%
  filter(date >= "2020-01-01")
```

Data summaries

What is the total number of outpatient visits in the dataset?

```
outpatient %>%  
  summarize(sum(outpatient_visits))
```

What is the mean (average) number of monthly outpatient visits?

```
outpatient %>%  
  summarize(mean(outpatient_visits))
```

What is the minimum and maximum number of monthly outpatient visits?

```
outpatient %>%  
  summarize(min(outpatient_visits),  
            max(outpatient_visits))
```

How many missing months are there?

```
outpatient %>%  
  summarize(sum(is.na(outpatient_visits)))
```

Creating a new variable

Create a new variable that indicates the month.

```
outpatient %>%  
  mutate(month = month(date))
```

Now, save the above as a new data frame (also known as a “tibble” in R)

```
outpatient %>%  
  group_by(month = month(date)) -> outpatient2  
  
head(outpatient2)
```

Summarize data within groupings

Over the time period, what is the average number of visits by month?

```
outpatient2 %>%  
  group_by(month) %>%  
  summarize(mean(outpatient_visits))
```

Lab activity

Please take 15 minutes to complete the following questions.

1. How many months have less than 3000 visits?
2. How many outpatient visits occurred in March 2019?
3. How many total outpatient visits occurred in 2016?
4. Load in the example_kmc.rds dataset
5. View the KMC dataset.
6. How many missing months are in this dataset?
7. What is the mean (average) number of monthly Kangaroo Mother Care?