# assignment 5

Michael Robinson

2023-10-06

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
# Load the required libraries
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)


Flight_info <- "https://raw.githubusercontent.com/MRobinson112/assignment-5/main/flightlab2.csv"

Airline_df <- read.csv(Flight_info, skip = 1, header = FALSE, stringsAsFactors = FALSE, na.strings = ""

# Removing blank rows
Airline_df <- Airline_df[complete.cases(Airline_df), ]

# Rename columns
colnames(Airline_df) <- c("Airlines", "Status", "Los_Angeles", "Phoenix", "San_Diego", "San_Francisco",

# Cleaning and transform the data
Clean_data <- Airline_df %>%
  pivot_longer(cols = c(Los_Angeles, Phoenix, San_Diego, San_Francisco, Seattle),
               names_to = "Destination",
               values_to = "Delay") %>%
```

```
  filter(!is.na(Delay)) %>%
  mutate(Status = ifelse(Status == "On time", "OnTime", "Delayed"))

Clean_data
```

```
## # A tibble: 10 x 4
##    Airlines  Status  Destination   Delay
##    <chr>     <chr>   <chr>         <int>
##  1 "ALASKA " Delayed Los_Angeles     497
##  2 "ALASKA " Delayed Phoenix         221
##  3 "ALASKA " Delayed San_Diego       212
##  4 "ALASKA " Delayed San_Francisco   503
##  5 "ALASKA " Delayed Seattle        1841
##  6 "AM WEST" Delayed Los_Angeles     694
##  7 "AM WEST" Delayed Phoenix        4840
##  8 "AM WEST" Delayed San_Diego       383
##  9 "AM WEST" Delayed San_Francisco   320
## 10 "AM WEST" Delayed Seattle         201
```

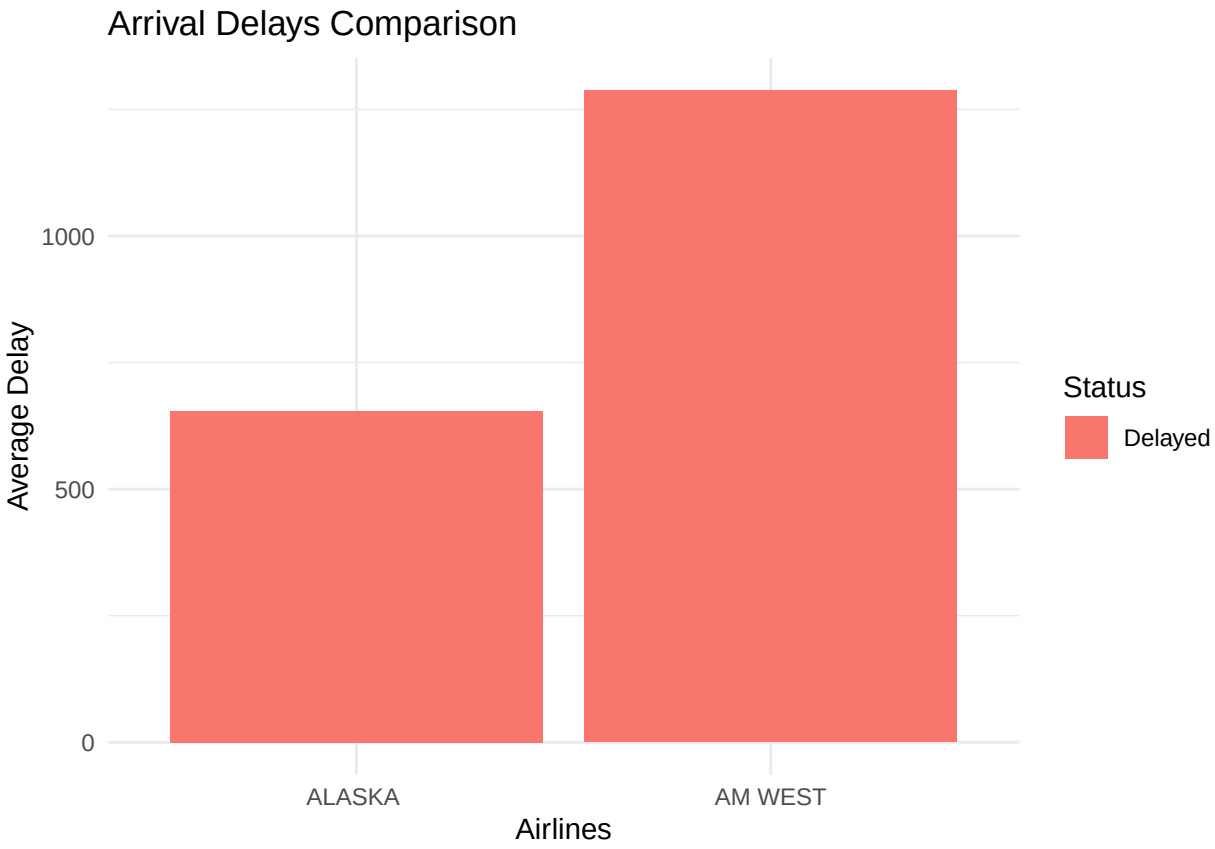
```
# Perform analysis
new_data <- Clean_data %>%
  group_by(Airlines, Status) %>%
  summarise(AvgDelay = mean(Delay), .groups = 'drop')
    new_data
```

```
## # A tibble: 2 x 3
##   Airlines  Status  AvgDelay
##   <chr>     <chr>      <dbl>
## 1 "ALASKA " Delayed     655.
## 2 "AM WEST" Delayed    1288.
```

```
delay_by_dest <- Clean_data %>%
  group_by(Destination, Status) %>%
  summarise(AvgDelay = mean(Delay), .groups = 'drop')
    delay_by_dest
```

```
## # A tibble: 5 x 3
##   Destination   Status  AvgDelay
##   <chr>         <chr>      <dbl>
## 1 Los_Angeles   Delayed     596.
## 2 Phoenix       Delayed    2530.
## 3 San_Diego     Delayed     298.
## 4 San_Francisco Delayed     412.
## 5 Seattle       Delayed    1021
```

```
# comparing arrival delays for the two airlines
ggplot(data = new_data, aes(x = Airlines, y = AvgDelay, fill = Status)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Arrival Delays Comparison",
       x = "Airlines",
       y = "Average Delay") +
  theme_minimal()
```


Arrival Delays Comparison
Average Delay
1000
500
0
ALASKA
AM WEST
Airlines
Status
Delayed