607_Assignment#5

Michael Robinson

2023-09-29

```
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)

# Read the data from the CSV file
flight_data <- read.csv("https://raw.githubusercontent.com/MRobinson112/assignment-5/main/flightdata.csv

# view the initial rows of the dataset.
head(flight_data)
```

```
##   Airline          City One_Time_Arrivals Delayed_Arrivals  X
## 1  Alaska   Los Angeles               497               62 NA
## 2  Alaska       Phoenix               221               12 NA
## 3  Alaska     San Diego               212               20 NA
## 4  Alaska San Francisco               503              102 NA
## 5  Alaska       Seattle              1841              305 NA
## 6 AM West   Los Angeles               694              117 NA
```

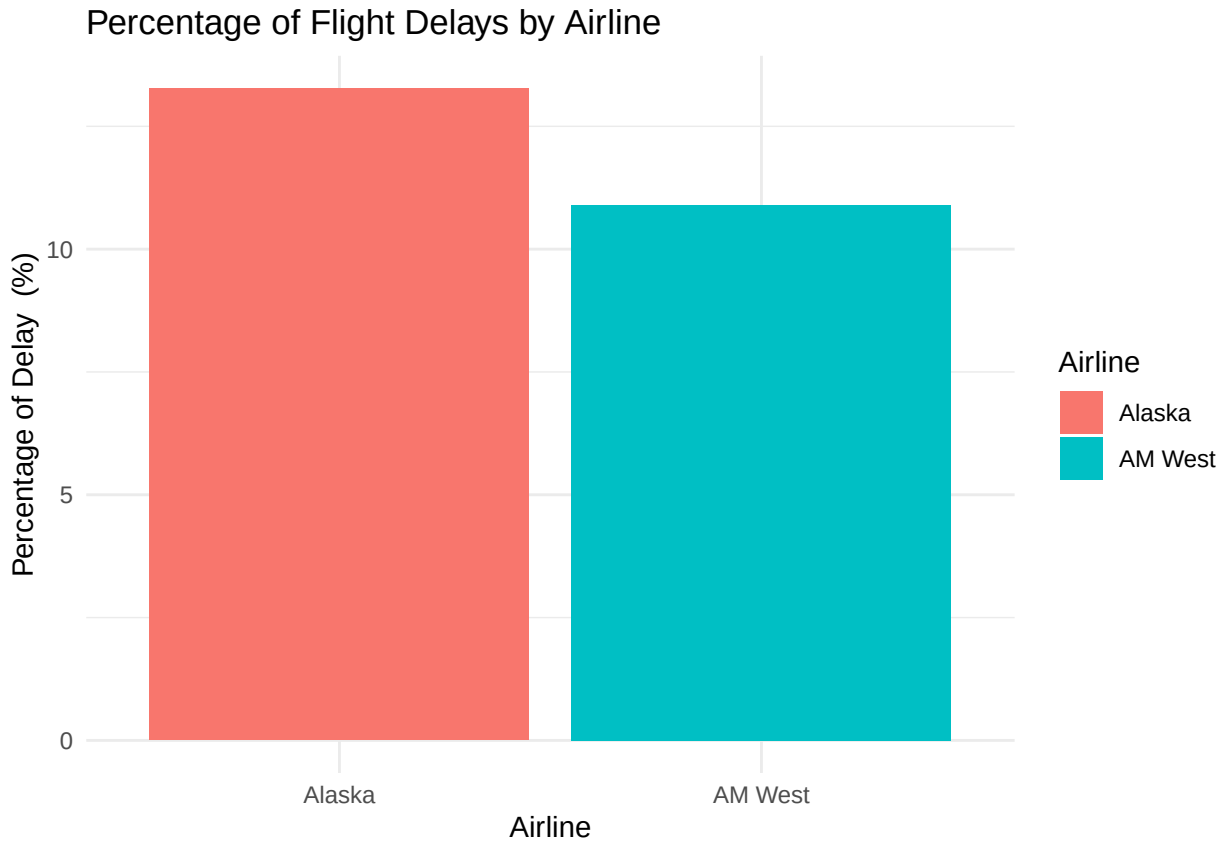
```
# Seperate data by airline and calculate the mean delay
delay_summary <- flight_data %>%
  group_by(Airline) %>%
  summarize(
   Average_Delay = mean(Delayed_Arrivals, na.rm = TRUE),
    Total_Flights = n()
  )

# View the summary
delay_summary
```

```
## # A tibble: 2 x 3
##   Airline Average_Delay Total_Flights
##   <chr>           <dbl>         <int>
## 1 AM West          157.             5
## 2 Alaska           100.             5
```

```r
# Determine the percentage of flights delay for each airline.
delay_percentages <- flight_data %>%
  group_by(Airline) %>%
  summarise(
    Percent_Delayed = (sum(Delayed_Arrivals) / sum(Delayed_Arrivals + One_Time_Arrivals)) * 100
  )

ggplot(delay_percentages, aes(x = Airline, y = Percent_Delayed, fill = Airline)) +
  geom_bar(stat = "identity") +
  labs(title = "Percentage of Flight Delays by Airline",
       x = "Airline",
       y = "Percentage of Delay  (%)") +
  theme_minimal()
```



Conclusion

Based on the summary data and the plot, it is evident that Alaska Airlines has the highest average percentage of flight delays among the airlines in the dataset.