Week 4 / Assignment – Tidying and Transforming Data

Alina Vikhnevich

2025-02-17

Introduction

This report focuses on tidying and transforming data in R. Using `tidyr` and `dplyr`, we'll reshape data between wide and long formats, making it more suitable for analysis.

Loading Data

```
flight_data <- read_csv("/Users/alina_vikhnevich/Desktop/Spring 2025/DATA 607/DATA607/flight_data.csv")
print(flight_data)
```

```
## # A tibble: 4 x 7
##   Airline Status  `Los Angeles` Phoenix `San Diego` `San Francisco` Seattle
##   <chr>   <chr>           <dbl>   <dbl>       <dbl>           <dbl>   <dbl>
## 1 ALASKA  on time           497     221         212             503    1841
## 2 ALASKA  delayed           62      12          20             102     305
## 3 AM WEST on time           694    4840         383             320     201
## 4 AM WEST delayed           117     415          65             129      61
```

Tidying the Data

Reshaping Wide to Long Format

```
tidy_flight_data <- flight_data %>%
  pivot_longer(cols = -c(Airline, Status),
               names_to = "Destination",
               values_to = "Count")
print(tidy_flight_data)
```

```
## # A tibble: 20 x 4
##    Airline Status  Destination   Count
##    <chr>   <chr>   <chr>         <dbl>
##  1 ALASKA  on time Los Angeles     497
##  2 ALASKA  on time Phoenix         221
##  3 ALASKA  on time San Diego       212
##  4 ALASKA  on time San Francisco   503
##  5 ALASKA  on time Seattle        1841
```

```
##  6 ALASKA  delayed Los Angeles      62
##  7 ALASKA  delayed Phoenix          12
##  8 ALASKA  delayed San Diego        20
##  9 ALASKA  delayed San Francisco   102
## 10 ALASKA  delayed Seattle         305
## 11 AM WEST on time Los Angeles     694
## 12 AM WEST on time Phoenix        4840
## 13 AM WEST on time San Diego       383
## 14 AM WEST on time San Francisco   320
## 15 AM WEST on time Seattle         201
## 16 AM WEST delayed Los Angeles     117
## 17 AM WEST delayed Phoenix         415
## 18 AM WEST delayed San Diego        65
## 19 AM WEST delayed San Francisco   129
## 20 AM WEST delayed Seattle          61
```

Summarizing the Data

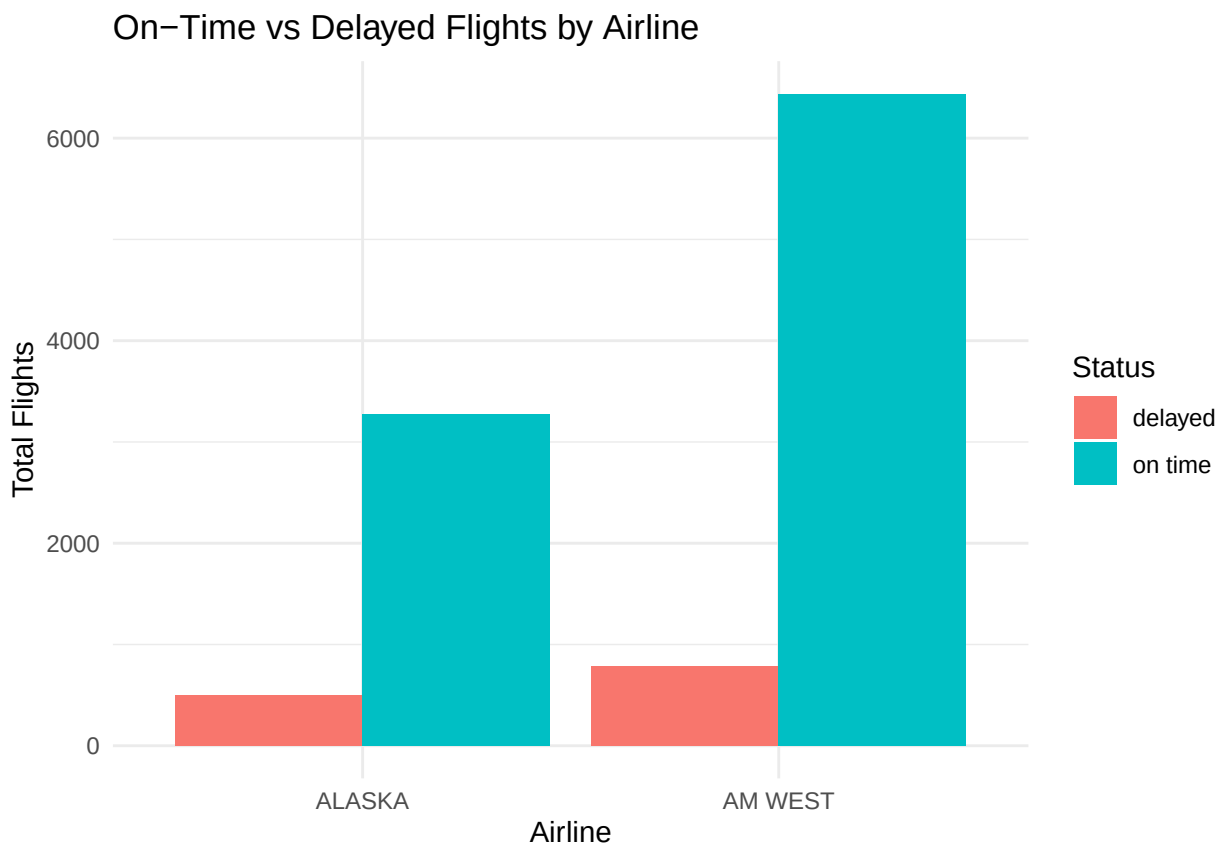
Total Flights by Airline and Status

```
summarized_data <- tidy_flight_data %>%
  group_by(Airline, Status) %>%
  summarise(Total_Flights = sum(Count), .groups = 'drop')
print(summarized_data)
```

```
## # A tibble: 4 x 3
##   Airline Status  Total_Flights
##   <chr>   <chr>           <dbl>
## 1 ALASKA  delayed           501
## 2 ALASKA  on time          3274
## 3 AM WEST delayed           787
## 4 AM WEST on time          6438
```

Visualizing the Data

Bar Plot: On-Time vs Delayed Flights



Conclusion

This assignment demonstrated how to reshape and analyze flight data in R. Using `tidyr` and `dplyr`, we converted wide-format data into a tidy structure and summarized key insights. This process makes data easier to work with for visualization and analysis.