

Visualization of flight arrival and departure pattern

Package and dataset needed

The `flight_df_raw` data we have here is a complete record of all domestic flights for American, Delta, United, and Southwest in 2019 (`nycflights13::flights` contains only flights departing from NYC).

```
library(tidyverse)
flight_df_raw <- read_csv("data/flight_df.csv")

# you can change the airline and airport later
airport_vec <- c("DFW", "CLT", "ORD", "AUS", "DEN", "SLC", "ATL", "PHL")
airline <- "AA"
```

Your Goal

Think of yourself as working for one of the big airline companies or Federal Aviation Administration. Your boss comes to you with this data and ask you to investigate the daily departure and arrival patterns for, say American Airlines (AA). They used to see the count of arrival/ departure flights shows a periodical cycle for some airports (below) and want to know if this pattern also exists for the airports we have above (DFW and CLT are AA hubs, PHL and ORD are relatively big airports, and there are some smaller ones: AUS, SLC, DEN, and ATL).

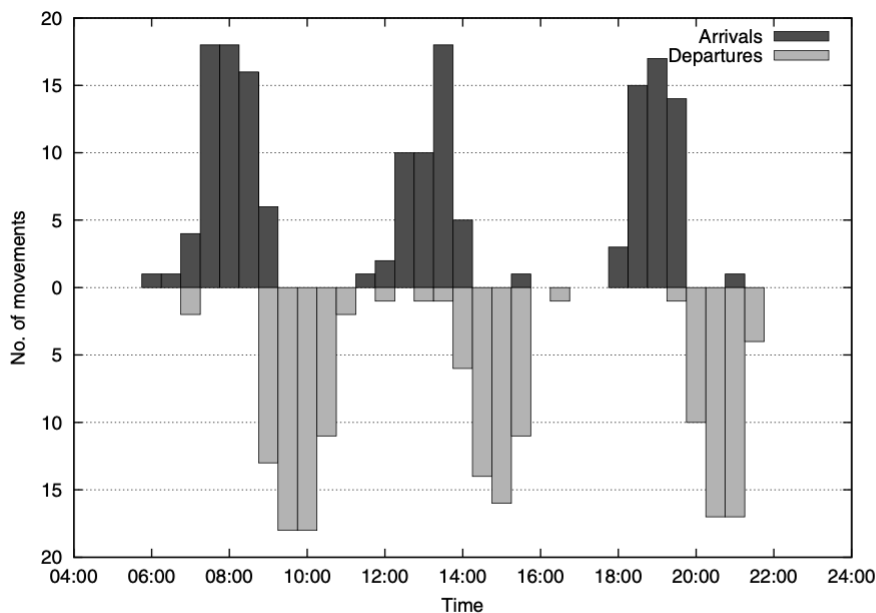


Fig. 2.4 Schedule structure of the Alitalia hub in Milan Malpensa (MXP) on 19th January 2005 (Source: Danesi (2006))

If you want to challenge yourself, you should try to start without hints and produce some visualizations to show your boss your investigation.

Hints

Here I provide a few instructions to decompose the task into smaller pieces, if you need some help.

- You may want to first narrow the data down to the airport and airline we want.
- The x-axis in the sample plot is the time of the day and this applies to both the departure time and arrival time.
 - The current `DepTime` and `ArrTime` are in character. Can you apply what we learn in **week 5 b date time** to convert them to a more useful format?
 - When creating the plot, we can map the arrival/ departure time to the x-axis if they are in the same column. Can you apply what we learn in data reshaping here?
- In the sample plot, each bar represents the count of flights in a 30-minute window. We will also need to do something like this to sort arrival and departure times into time intervals.
 - The `lubridate` package has the function `floor_date()` that allows you to round down a date-time object to the nearest unit you want (e.g., 10 minutes, 2 hours).
 - You may want to play around with the window size (e.g., 10 minutes, 20 minutes, 1 hour) to see how it affects the visualization.
 - You will need to count the number of flights that falls into each time window.
- Now you're ready to create the plot - when constructing the plot, think about the following:
 - What aesthetic mappings do you need?
 - Do you need to facet the plot?
 - What would you do to show arrival and departure type as the sample plot?

```
flight_df_raw
```

```
## # A tibble: 3,119,757 x 6
##   Reporting_Airline FlightDate DepTime ArrTime Origin Dest
##   <chr>             <date>     <chr>   <chr>   <chr> <chr>
## 1 AA               2019-01-01 0003     0504   LAX    DFW
## 2 AA               2019-01-01 0003     0150   CLT    PBI
## 3 AA               2019-01-01 0004     0912   OGG    PHX
## 4 AA               2019-01-01 0005     0723   LAX    ATL
## 5 AA               2019-01-01 0011     0832   SFO    MIA
## 6 AA               2019-01-01 0014     0224   PHX    SEA
## 7 AA               2019-01-01 0014     0543   PHX    CLT
## 8 AA               2019-01-01 0015     0333   DFW    MCO
## 9 AA               2019-01-01 0015     0346   DFW    MIA
## 10 AA              2019-01-01 0015     0108   PHX    RNO
## # i 3,119,747 more rows
```