Data on air traffic at DEN and ATL

Bhavya Bhargava

March 22, 2024

Importing the packages that were utilized and citing them

```
library(tidyverse)
citation("tidyverse")
```

```
## To cite package 'tidyverse' in publications use:
##
##   Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R,
##   Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller
##   E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V,
##   Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to
##   the tidyverse." _Journal of Open Source Software_, *4*(43), 1686.
##   doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {Welcome to the {tidyverse}},
##     author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Luc
y D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henr
y and Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bach
e and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spin
u and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani},
##     year = {2019},
##     journal = {Journal of Open Source Software},
##     volume = {4},
##     number = {43},
##     pages = {1686},
##     doi = {10.21105/joss.01686},
##   }
```

How the data was downloaded, and processed

Data was downloaded from the "transtats.bts.gov" website and the path to get the below data was:

1. Went to the aviation section in the side menu.

2. Under the Databases section selected the "Airline On-Time Performance data"

3. Under this databases page selected the "Reporting Carrier On-Time Performance (1987-present)" option.

4. On the new page which opened selected the download option from the side menu.

On the download page with set parameters for timeline and fields tried downloading the dataset, but except the time constraints no filtering took place from the website and the whole un-filtered data set was downloaded.

The downloaded dataset has been placed in the data directory for this markdown and all it's filtering and processing will happen below for a smaller subset and subsequent resulting .rda file:

```
# Reading the CSV file into flights_data(Perhaps changing the below path to just "data/On_Time_Performance_Data.csv" might help.)
flights_data <- read_csv(here::here("C:/Users/user/Desktop/ETC5512/assignment1/data/On_Time_Performance_Data.csv"))

# Filtering fields according to the selected airports and required data
selected_airports_flights <- flights_data %>%
  filter((Origin %in% c("DEN", "ATL") | Dest %in% c("DEN", "ATL"))) %>%
  select(DayOfWeek, FlightDate, Reporting_Airline,
         DOT_ID_Reporting_Airline, IATA_CODE_Reporting_Airline,
         Origin, Dest, OriginCityName, DestCityName,
         OriginState, DestState, DepDelay, DepDel15,
         ArrDelay, ArrDel15, Cancelled, CancellationCode,
         Diverted, Flights, Distance, CRSElapsedTime, ActualElapsedTime)

# Filtered data into .rda file (the directories in the read and save part are same (please modify if needed))
save(selected_airports_flights, file = "data/airports.rda")
```

This statement above gives us the "airports.rda" and saves it in the data folder.

Dataset description

The data is in file `airports.rda` in the `data` directory. It contains these variables:

- Time Period DayOfWeek and FlightDate -> These fields have been selected as they might be useful to Qantas in identifying which days and dates the traffic and demand is most helping them planning their flight schedules.

- Airline Reporting_Airline, DOT_ID_Reporting_Airline and IATA_CODE_Reporting_Airline -> As Qantas doesn't just operate it's own flights but many codeshare flights along with it's subsidiaries these fields will be useful in identifying those.

- Origin and Destination

–Origin, Dest -> Filtering of flights to and from ATL and DEN was supported by these fields

–OriginCityName, DestCityName -> As many a times the name of the city being travelled to or from is not that obvious from their abbreviations thus to give a high level idea on their names these fileds were selected.

–OriginState, DestState -> Useful for getting a broad overview of connectivity of the airport across the country.

- Departure and Arrival Performance

–DepDelay and ArrDelay -> These fields are very useful as they give the exact amount of minutes the flight departed or arrived early or late.These fields also help assess the reliability of departures and arrivals from DEN and ATL, which is crucial for minimizing missed connections and other accidents.

- Cancellations and Diversions

–Cancelled, CancellationCode -> Important for understanding the frequency and reasons behind cancellations at our selected airports, as high cancellation rates can significantly disrupt connectivity and cause further troubles.

–Diverted -> This gives insight into how often flights scheduled for these airports are diverted, another factor in assessing their reliability.

- Flight Summaries

–Flights, Distance -> These fields may prove useful in providing insights into the amount of flights and the range of destinations covered, important for evaluating the extent of connectivity.

–CRSElapsedTime, ActualElapsedTime -> Helps assess the different calculation of duration of flights, offering indirect insights on reachability of potential destinations.

Dataset population and limitations analysis

In case of this dataset, the larger population includes all regularly scheduled flights in the United States, with a focus on those that connect via DEN and ATL throughout the year. This comprehensive view helps you understand how these airports connect to air traffic networks across the country and around the world. Speaking of the limitations of the data, as it covers only seasonal data it might not show the full patterns for the year including passenger volumes and operational delays.

Dataset licensing situation

This dataset qualifies as open data as it adheres to the principles set by the U.S. open data policy. Under this policy, data released by the U.S. Federal government through Data.gov is made available for public access and use without any restrictions. (https://data.gov/privacy-policy/#licensing (https://data.gov/privacy-policy/#licensing))

To understand the created subset and understand which airport has the better connectivity

```r
# Loading the data subset for analysis(Perhaps changing the below path to just "data/airport
s.rda" might help.)
load(here::here("C:/Users/user/Desktop/ETC5512/assignment1/data/airports.rda"))

# Counting unique states connectivity to ATL Airport by filtering (excluding GA)
atl_flights <- selected_airports_flights %>%
   filter(Origin == "ATL" | Dest == "ATL") %>%
   select(Origin, Dest, OriginState, DestState)

# Count unique Origin States
atl_unique_origin_states <- atl_flights %>%
   filter(OriginState != "GA") %>%
   distinct(OriginState) %>%
   nrow()

# Count unique Dest States
atl_unique_dest_states <- atl_flights %>%
   filter(DestState != "GA") %>%
   distinct(DestState) %>%
   nrow()

# Counting unique states connectivity to DEN Airport by filtering (excluding CO)
den_flights <- selected_airports_flights %>%
   filter(Origin == "DEN" | Dest == "DEN") %>%
   select(Origin, Dest, OriginState, DestState)

# Count unique Origin States
den_unique_origin_states <- den_flights %>%
   filter(OriginState != "CO") %>%
   distinct(OriginState) %>%
   nrow()

# Count unique Dest States
den_unique_dest_states <- den_flights %>%
   filter(DestState != "CO") %>%
   distinct(DestState) %>%
   nrow()

# Presenting Results for Comparision
cat("Unique Origin States for ATL (excluding GA):", atl_unique_origin_states, "\n")
```

```
## Unique Origin States for ATL (excluding GA): 47
```

```r
cat("Unique Destination States for ATL (excluding GA):", atl_unique_dest_states, "\n")
```

```
## Unique Destination States for ATL (excluding GA): 47
```

```r
cat("Unique Origin States for DEN (excluding CO):", den_unique_origin_states, "\n")
```

```
## Unique Origin States for DEN (excluding CO): 43
```

```
cat("Unique Destination States for DEN (excluding CO):", den_unique_dest_states, "\n")
```

```
## Unique Destination States for DEN (excluding CO): 43
```

As conclusive from the above observations, the Airport in Atlanta i.e. ATL has better connectivity than the one in Denver and would be the best in terms of providing connecting flights to other parts of the country.