ERAU Capstone

# Introduction

This document was produced in RStudio, and it is currently a work in progress for the ERAU Capstone course. The main objective is to provide a detailed, step-by-step description of the data collection, preparation, and analysis processes, as agreed on the proposal.

Besides, it aims to allow traceability and reproducibility of the entire research, for potential future developments and academic discussion.

The first lines below will install and load packages necessary to further steps.

#Installing packages  
#install.packages("dplyr")  
#install.packages("ggstatsplot")  
#install.packages("lubridate")  
#install.packages("tidyverse")  
#install.packages("ggpubr")  
#install.packages("PairedData")  
  
#Loading packages  
library(dplyr)  
library(lubridate)  
library(tidyverse)  
library(ggstatsplot)  
library(rmarkdown)  
library(ggpubr)  
library(PairedData)

## Listing useful filters

The code below lists all the initial letters from Brazilian airports ICAO codes, according to DECEA. They will be useful on the future to filter the data accordingly.

bra\_prefixes <- c("^SB", "^SD", "^SI", "^SJ", "^SN", "^SS", "^SW", "^9P", "^ZZ")

# Data Collection

Data was downloaded from <https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/historico-de-voos> . The website provides monthly csv files, that were saved in a folder “/ANAC\_DATABASE”, and renamed to a standardized form such as: “vra\_01\_2019.csv”; “vra\_02\_2019.csv”; “vra\_03\_2019” and so on up until “vra\_12\_2020.csv”.

# Data Preparation

Before starting the analysis, data needs to be prepared in order to organize the variable names, clean invalid entries, delimit the geographical scope and other adjustments.

The following scripts were used to read all files and create the base dataset files “vra\_full\_2019.csv” and “vra\_full\_2020.csv”. They were written on separate R files for easier reading of this document. As they are already executed, and the base csv files are already saved, both lines were disabled (commented with #)

#source("./R/data/cleaning\_anac\_2019.R")  
#source("./R/data/cleaning\_anac\_2020.R")

Finally, both csv files were loaded into the current environment, as dataframes “vra\_2019” and “vra\_2020”, which allows to operate them. The variable names and types (categorical x continous) were also adjusted.

Below, the reader can find a summary of 2019 dataset. ANAC database has 982991 entries, including realized and canceled flights, but still many invalid entries (“NAs”).

glimpse(vra\_2019)

## Rows: 982,991  
## Columns: 12  
## $ ICAO\_Airline <fct> AAF, AAF, AAF, AAF, AAF, AAF, AAF, AAF, AAL, AAL...  
## $ Flight\_number <fct> AAF35, AAF35, AAF35, AAF36, AAF36, AAF36, AAF37,...  
## $ Authorization\_type <fct> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## $ Operation\_type <fct> I, I, I, I, I, I, I, I, I, I, I, I, I, I, I, I, ...  
## $ ADEP <fct> LFPO, LFPO, LFPO, SBKP, SBKP, SBKP, LFPO, SBKP, ...  
## $ ADES <fct> SBKP, SBKP, SBKP, LFPO, LFPO, LFPO, SBKP, LFPO, ...  
## $ SOBT <dttm> 2019-01-25 06:15:00, 2019-01-27 06:15:00, 2019-...  
## $ AOBT <dttm> 2019-01-25 06:15:00, 2019-01-27 06:15:00, 2019-...  
## $ SIBT <dttm> 2019-01-25 18:15:00, 2019-01-27 18:15:00, 2019-...  
## $ AIBT <dttm> 2019-01-25 18:15:00, 2019-01-27 18:15:00, 2019-...  
## $ Status <fct> REALIZADO, REALIZADO, REALIZADO, REALIZADO, REAL...  
## $ Justification\_code <fct> NA, NA, NA, NA, NA, NA, NA, NA, AT, NA, RA, TD, ...

Finally, a summary of the 2020 dataset. ANAC database has 485166 entries, including realized and canceled flights, but still many invalid entries (“NAs”).

glimpse(vra\_2020)

## Rows: 485,166  
## Columns: 12  
## $ ICAO\_Airline <fct> TAM, TAM, TAM, TAM, TAM, TAM, TAM, TAM, TAM, TAM...  
## $ Flight\_number <fct> TAM9458, TAM9201, TAM9200, TAM9046, TAM9010, TAM...  
## $ Authorization\_type <fct> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## $ Operation\_type <fct> I, N, N, N, N, N, N, N, N, N, N, N, I, I, I, I, ...  
## $ ADEP <fct> SBGR, SBSG, SBGL, SBGR, SBMO, SBGR, SBSG, SBGR, ...  
## $ ADES <fct> SCEL, SBGL, SBSG, SBBR, SBGR, SBMO, SBGR, SBSG, ...  
## $ SOBT <dttm> 2020-01-01 06:30:00, 2020-01-01 02:00:00, 2020-...  
## $ AOBT <dttm> 2020-01-01 06:30:00, 2020-01-01 02:00:00, 2020-...  
## $ SIBT <dttm> 2020-01-01 10:40:00, 2020-01-01 05:10:00, 2020-...  
## $ AIBT <dttm> 2020-01-01 10:40:00, 2020-01-01 05:10:00, 2020-...  
## $ Status <fct> REALIZADO, REALIZADO, REALIZADO, REALIZADO, REAL...  
## $ Justification\_code <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...

## Filtering invalid data, and only realized flights

The code below filters the observations for which the data entry is invalid, e.g. “NA”. In addition, it also removes canceled and not executed flights.

filter\_na <- . %>% filter(!is.na(ADEP) & !is.na(ADES) & !is.na(SOBT) & !is.na(AOBT) & !is.na(SIBT) & !is.na(AIBT)) %>% filter(Status == "REALIZADO")  
  
vra\_2019\_realizado <- vra\_2019 %>% filter\_na()  
vra\_2020\_realizado <- vra\_2020 %>% filter\_na()

### Check the size of the datasets after data filtering

Up until now, after removing invalid entries, it is important to check the sample size.

The 2019 sample has 982991 flights, from which 965183 (98.1883863%) were realized.

The 2020 sample has 485166 flights, from which 459578 (94.7259289%) were realized.

# Treatment of Outliers and Filtering Brazilian Data

Preliminarily, it is possible to identify some sources of outliers. To treat then, some assumptions were made. The work will remove:

* Local flights (i.e. flights with the same origin and destination)
* Arrival delays lower than -6 hours and higher than 6 hours
* Flight times less than 15 minutes or higher than 12 hours

## Removing local flights

The code below will remove local flights from the dataset.

no\_local <- . %>% filter(as.character(ADEP) != as.character(ADES))  
  
vra\_2019\_realizado\_loc <- vra\_2019\_realizado %>% no\_local()  
vra\_2020\_realizado\_loc <- vra\_2020\_realizado %>% no\_local()

### Check the size of the datasets after removing local flights

Up until now, 2019 sample has 982991 flights, from which 962311 (97.8962168%) were realized, non-locals. This is the base 2019 sample.

Up until now, 2020 sample has 485166 flights, from which 458870 (94.5799994%) were realized, non-locals. This is the base 2020 sample.

## Regarding the arrival delays branch of the research

The code below creates a reduced table with only the relevant data for delay analysis - airport, year, and arrival delay.

Relevant assumptions:

* Early arrivals (Delay < 0) must be adjusted to 0 minutes of delay
* Only arrival delays between -6 and 6 hours.

As a result, 2019 delays dataset was reduced to this:

glimpse(arrdly\_2019)

## Rows: 875,794  
## Columns: 4  
## $ ADES <fct> SBKP, SBKP, SBKP, SBKP, SBBR, SBBR, SBBR, SBBR, SBBR, S...  
## $ YR <fct> 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2...  
## $ ARR\_DLY <dbl> 0, 0, 0, 0, -32, 0, 142, 0, 0, 0, 0, 0, 0, 0, -43, -36,...  
## $ ARR\_DLY\_ADJ <dbl> 0, 0, 0, 0, 0, 0, 142, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...

And 2020 delays dataset was reduced to this:

glimpse(arrdly\_2020)

## Rows: 426,552  
## Columns: 4  
## $ ADES <fct> SBGL, SBSG, SBBR, SBGR, SBMO, SBGR, SBSG, SBBR, SBSG, S...  
## $ YR <fct> 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2...  
## $ ARR\_DLY <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...  
## $ ARR\_DLY\_ADJ <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...

## Regarding the flight times branch of the research

The code below creates a reduced table with only the relevant data for flight time analysis - departure airport, destination airport, year, and flight duration. Relevant assumptions:

* Only flights with Brazilian airports at origin or destination
* Only flights with duration between 15 minutes and 12 hours
* No local flights

As a result, 2019 flight duration dataset was reduced to this:

glimpse(flttimes\_2019)

## Rows: 940,234  
## Columns: 4  
## $ ADEP <fct> LFPO, LFPO, LFPO, SBKP, SBKP, SBKP, LFPO, SBKP, KMIA, KMIA...  
## $ ADES <fct> SBKP, SBKP, SBKP, LFPO, LFPO, LFPO, SBKP, LFPO, SBBR, SBBR...  
## $ YR <fct> 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2019...  
## $ FLT\_TIME <dbl> 12.000000, 12.000000, 12.000000, 11.500000, 11.500000, 11....

And 2020 flight duration dataset was reduced to this:

glimpse(flttimes\_2020)

## Rows: 449,974  
## Columns: 4  
## $ ADEP <fct> SBGR, SBSG, SBGL, SBGR, SBMO, SBGR, SBSG, SBGR, SBJP, SBBR...  
## $ ADES <fct> SCEL, SBGL, SBSG, SBBR, SBGR, SBMO, SBGR, SBSG, SBBR, SBSG...  
## $ YR <fct> 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020...  
## $ FLT\_TIME <dbl> 4.166667, 3.166667, 3.083333, 1.833333, 3.166667, 2.916667...

# Data Transformation to Proper Research Tables

Finally, both delay and flight duration datasets must be transformed to the proposal’s approved format.

## Airports mean arrival delays

Here, another assumption must be made to prevent outliers:

* Only airports with 10 or more arrivals

## # A tibble: 146 x 3  
## AIRPORT AVG\_DELAY\_2019 AVG\_DELAY\_2020  
## <fct> <dbl> <dbl>  
## 1 SBGR 8.18 6.61  
## 2 SBSP 6.90 4.61  
## 3 SBBR 6.08 4.75  
## 4 SBKP 8.72 7.69  
## 5 SBCF 6.82 5.89  
## 6 SBGL 7.34 5.98  
## 7 SBRJ 6.24 4.31  
## 8 SBRF 7.19 5.38  
## 9 SBPA 6.64 5.11  
## 10 SBCT 6.85 6.04  
## # ... with 136 more rows

As a result, the final table is ready for analysis, containing 146 airports.

mean\_delays

## # A tibble: 146 x 3  
## AIRPORT AVG\_DELAY\_2019 AVG\_DELAY\_2020  
## <fct> <dbl> <dbl>  
## 1 SBGR 8.18 6.61  
## 2 SBSP 6.90 4.61  
## 3 SBBR 6.08 4.75  
## 4 SBKP 8.72 7.69  
## 5 SBCF 6.82 5.89  
## 6 SBGL 7.34 5.98  
## 7 SBRJ 6.24 4.31  
## 8 SBRF 7.19 5.38  
## 9 SBPA 6.64 5.11  
## 10 SBCT 6.85 6.04  
## # ... with 136 more rows

## Average flight time between city pairs

Here, another assumption must be made to prevent outliers:

* Only routes with 730 or more flights (around 2 flights per day in average)

As a result the flight times table is presented below:

avg\_city\_pairs

## # A tibble: 327 x 5  
## CITY\_PAIR FLTS\_2019 AVG\_FLT\_TIME\_2019 FLTS\_2020 AVG\_FLT\_TIME\_2020  
## <chr> <int> <dbl> <int> <dbl>  
## 1 SBSP-SBRJ 17671 0.986 7316 0.979  
## 2 SBRJ-SBSP 17616 1.05 7298 1.05   
## 3 SBPA-SBGR 8416 1.71 4390 1.65   
## 4 SBGR-SBPA 8377 1.73 4406 1.71   
## 5 SBSP-SBBR 7771 1.71 2899 1.69   
## 6 SBBR-SBSP 7758 1.70 2894 1.67   
## 7 SBCF-SBSP 7725 1.30 3017 1.29   
## 8 SBSP-SBCF 7655 1.22 3008 1.20   
## 9 SBSP-SBPA 6917 1.64 2282 1.61   
## 10 SBPA-SBSP 6864 1.54 2260 1.50   
## # ... with 317 more rows

summary(avg\_city\_pairs)

## CITY\_PAIR FLTS\_2019 AVG\_FLT\_TIME\_2019 FLTS\_2020   
## Length:327 Min. : 738 Min. : 0.5562 Min. : 114.0   
## Class :character 1st Qu.: 1155 1st Qu.: 1.2271 1st Qu.: 463.5   
## Mode :character Median : 1533 Median : 1.6142 Median : 756.0   
## Mean : 2184 Mean : 2.2741 Mean :1021.7   
## 3rd Qu.: 2502 3rd Qu.: 2.6130 3rd Qu.:1176.0   
## Max. :17671 Max. :11.6768 Max. :7316.0   
## AVG\_FLT\_TIME\_2020  
## Min. : 0.5453   
## 1st Qu.: 1.2083   
## Median : 1.6058   
## Mean : 2.2608   
## 3rd Qu.: 2.6364   
## Max. :11.6637

However, it is still necessary to normalize the 2020 values to 2019 references.

As a result, the final table for analysis is presented below, with 327 routes :

## # A tibble: 327 x 3  
## CITY\_PAIR NORM\_FLT\_2020 NORM\_FLT\_TIME\_2020  
## <chr> <dbl> <dbl>  
## 1 SBSP-SBRJ 0.414 0.993  
## 2 SBRJ-SBSP 0.414 1.00   
## 3 SBPA-SBGR 0.522 0.969  
## 4 SBGR-SBPA 0.526 0.984  
## 5 SBSP-SBBR 0.373 0.992  
## 6 SBBR-SBSP 0.373 0.977  
## 7 SBCF-SBSP 0.391 0.988  
## 8 SBSP-SBCF 0.393 0.988  
## 9 SBSP-SBPA 0.330 0.981  
## 10 SBPA-SBSP 0.329 0.974  
## # ... with 317 more rows

# Data Analysis

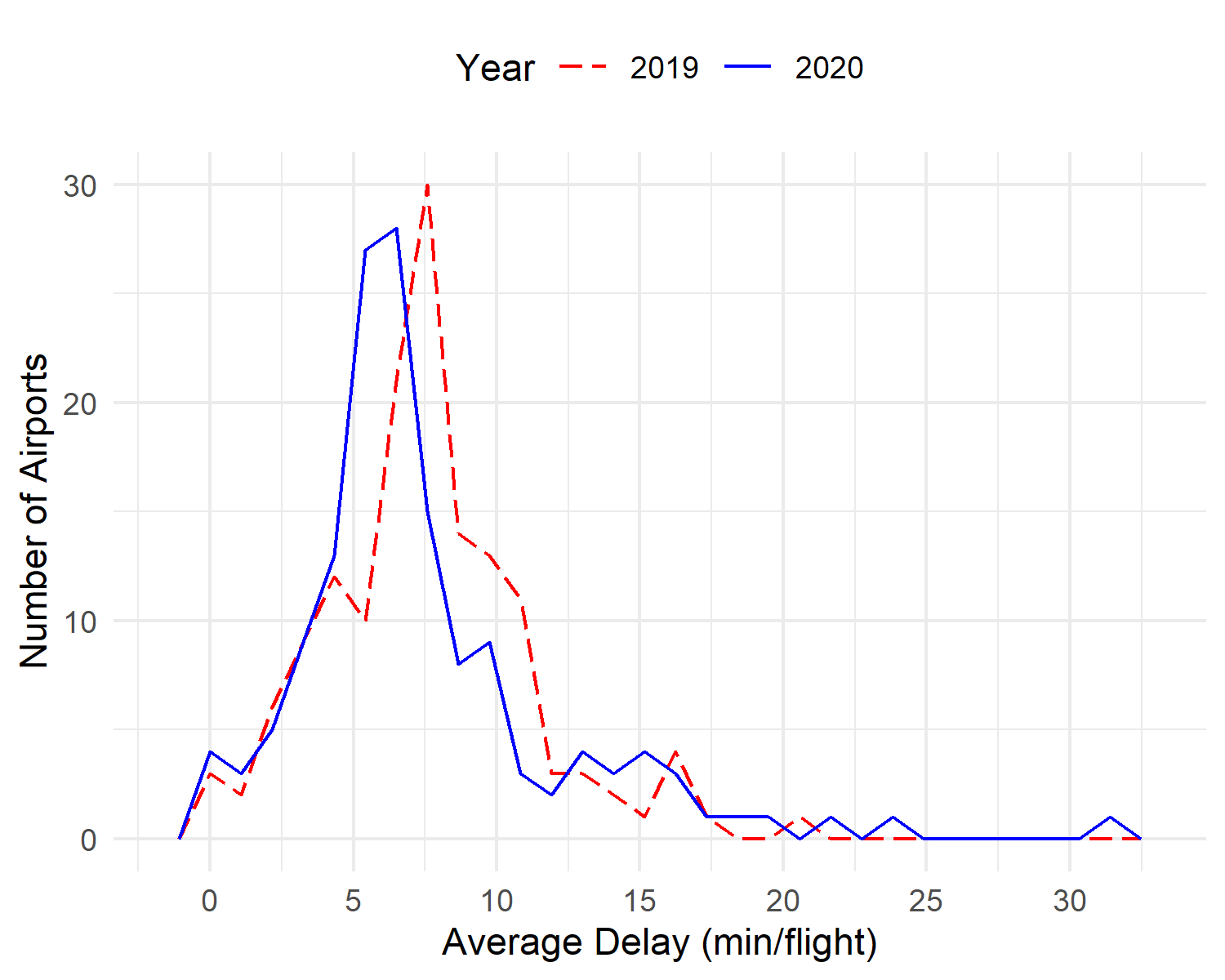
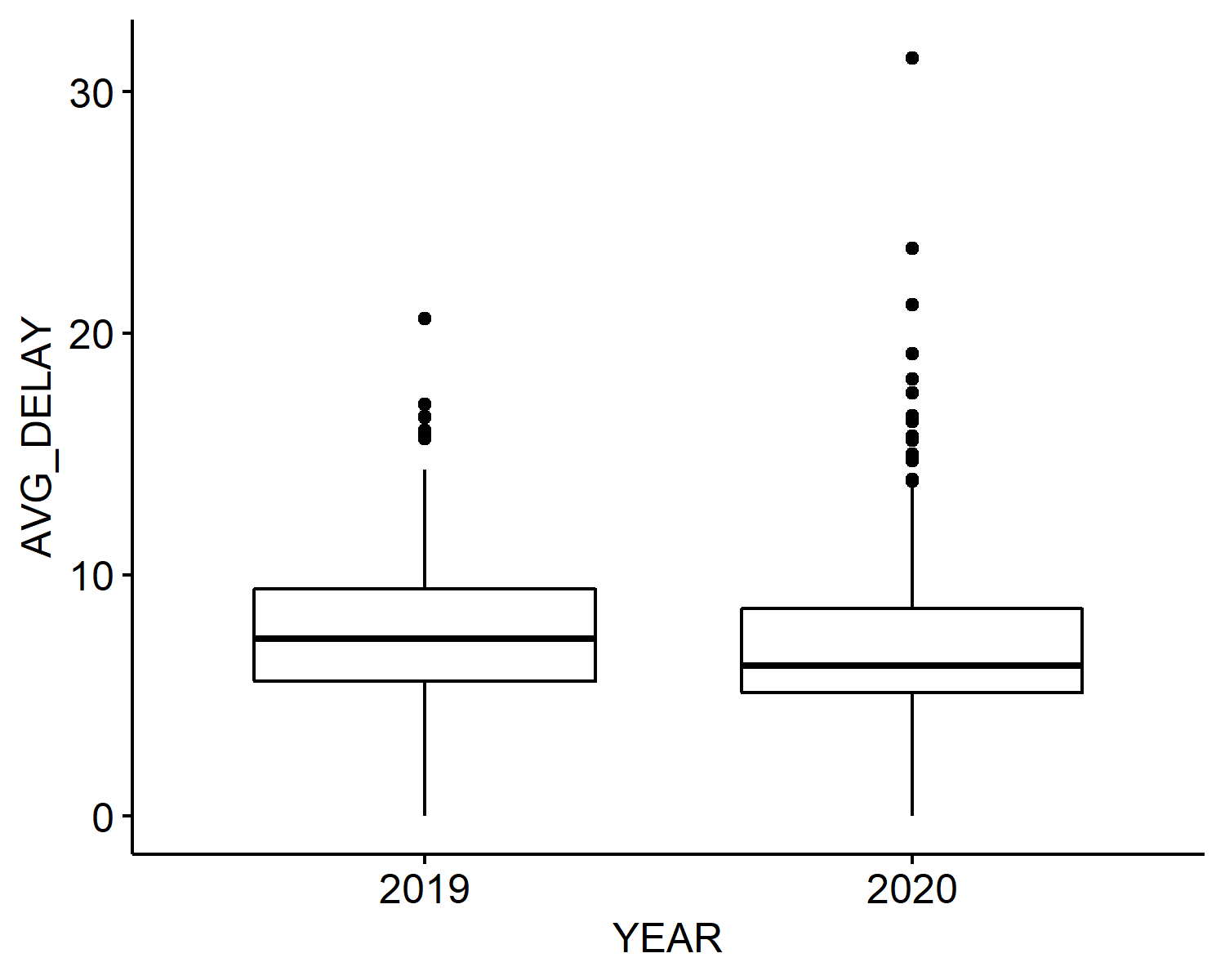
## Arrival delays paired T-test

## # A tibble: 10 x 7  
## AIRPORT ARRIVALS\_2019 AVG\_DELAY\_2019 ARRIVALS\_2020 AVG\_DELAY\_2020  
## <fct> <int> <dbl> <int> <dbl>  
## 1 SBGR 136214 8.18 72336 6.61  
## 2 SBSP 87161 6.90 29852 4.61  
## 3 SBBR 60229 6.08 30855 4.75  
## 4 SBKP 55956 8.72 34443 7.69  
## 5 SBCF 49512 6.82 22158 5.89  
## 6 SBGL 47001 7.34 16512 5.98  
## 7 SBRJ 41753 6.24 22884 4.31  
## 8 SBRF 35325 7.19 19670 5.38  
## 9 SBPA 33333 6.64 14633 5.11  
## 10 SBCT 30445 6.85 12152 6.04  
## # ... with 2 more variables: DECREASE\_ARR <lgl>, DECREASE\_DEL <lgl>

## # A tibble: 10 x 7  
## AIRPORT ARRIVALS\_2019 AVG\_DELAY\_2019 ARRIVALS\_2020 AVG\_DELAY\_2020  
## <fct> <int> <dbl> <int> <dbl>  
## 1 SSFB 38 4.87 33 7.42  
## 2 SSGY 38 7.79 32 6.03  
## 3 SSPI 37 13.4 34 16.6   
## 4 SJRG 23 0 62 4.47  
## 5 SSSB 22 16.5 42 23.5   
## 6 SDIY 18 2.44 12 3.67  
## 7 SBAQ 8 0 45 23.6   
## 8 SBGU 2 0 3 0   
## 9 SBNT 1 0 3 0   
## 10 SSGG 1 13 7 6.43  
## # ... with 2 more variables: DECREASE\_ARR <lgl>, DECREASE\_DEL <lgl>

## AIRPORT ARRIVALS\_2019 AVG\_DELAY\_2019 ARRIVALS\_2020   
## SBAC : 1 Min. : 1.0 Min. : 0.000 Min. : 1.00   
## SBAE : 1 1st Qu.: 123.0 1st Qu.: 5.518 1st Qu.: 67.25   
## SBAQ : 1 Median : 372.5 Median : 7.335 Median : 238.50   
## SBAR : 1 Mean : 5820.5 Mean : 7.471 Mean : 2838.03   
## SBAT : 1 3rd Qu.: 2439.2 3rd Qu.: 9.400 3rd Qu.: 1200.00   
## SBAU : 1 Max. :136214.0 Max. :20.620 Max. :72336.00   
## (Other):144   
## AVG\_DELAY\_2020 DECREASE\_ARR DECREASE\_DEL   
## Min. : 0.000 Mode :logical Mode :logical   
## 1st Qu.: 5.055 FALSE:24 FALSE:55   
## Median : 6.224 TRUE :126 TRUE :95   
## Mean : 7.461   
## 3rd Qu.: 8.606   
## Max. :31.397   
##

## # A tibble: 2 x 5  
## YEAR count mean median sd  
## <chr> <int> <dbl> <dbl> <dbl>  
## 1 2019 146 7.59 7.35 3.55  
## 2 2020 146 7.46 6.22 4.62



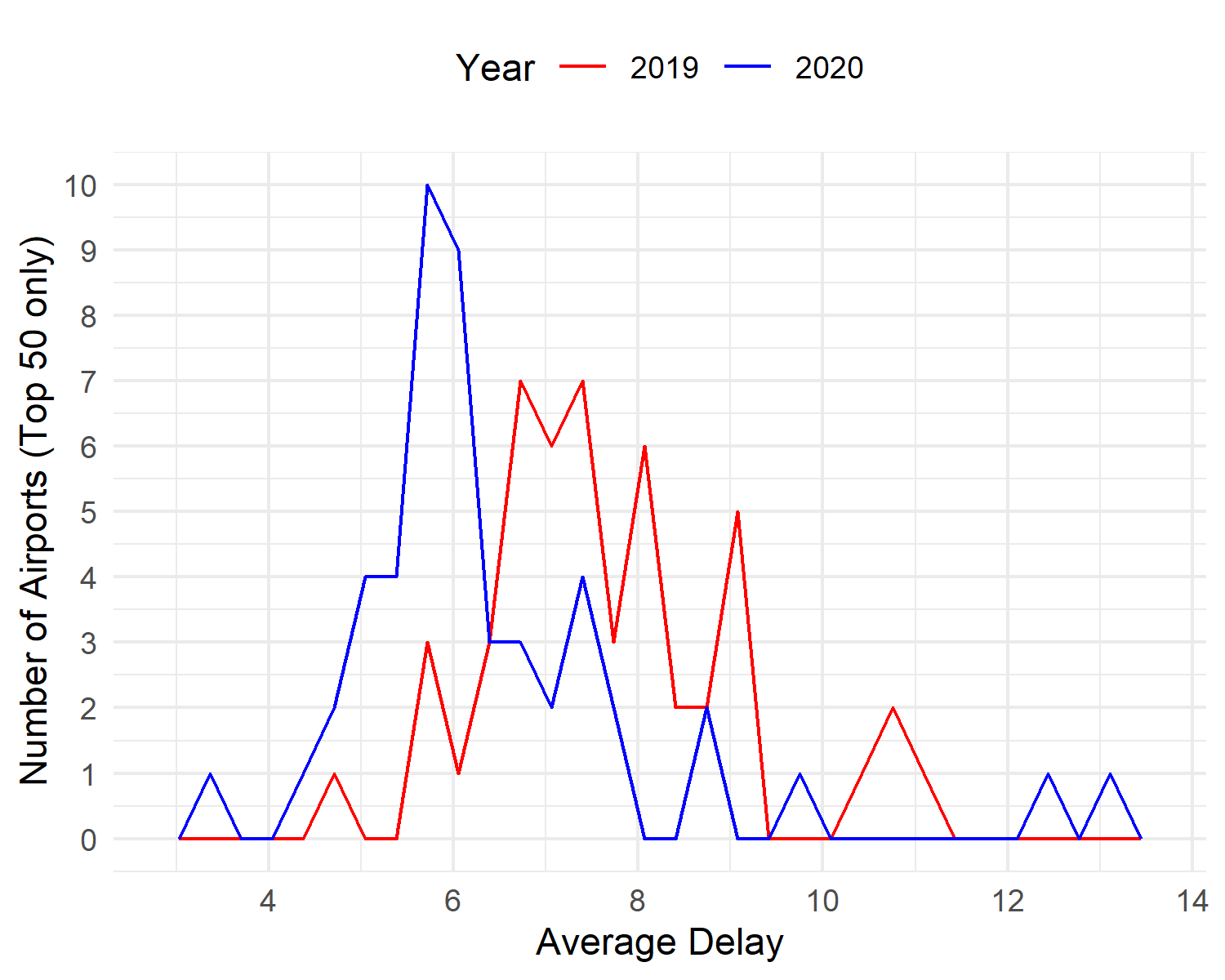
##   
## Paired t-test  
##   
## data: AVG\_DELAY by YEAR  
## t = 0.38222, df = 145, p-value = 0.7029  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.5297815 0.7838093  
## sample estimates:  
## mean of the differences   
## 0.1270139

## # A tibble: 50 x 5  
## AIRPORT ARRIVALS\_2019 AVG\_DELAY\_2019 ARRIVALS\_2020 AVG\_DELAY\_2020  
## <fct> <int> <dbl> <int> <dbl>  
## 1 SBGR 136214 8.18 72336 6.61  
## 2 SBSP 87161 6.90 29852 4.61  
## 3 SBBR 60229 6.08 30855 4.75  
## 4 SBKP 55956 8.72 34443 7.69  
## 5 SBCF 49512 6.82 22158 5.89  
## 6 SBGL 47001 7.34 16512 5.98  
## 7 SBRJ 41753 6.24 22884 4.31  
## 8 SBRF 35325 7.19 19670 5.38  
## 9 SBPA 33333 6.64 14633 5.11  
## 10 SBCT 30445 6.85 12152 6.04  
## # ... with 40 more rows

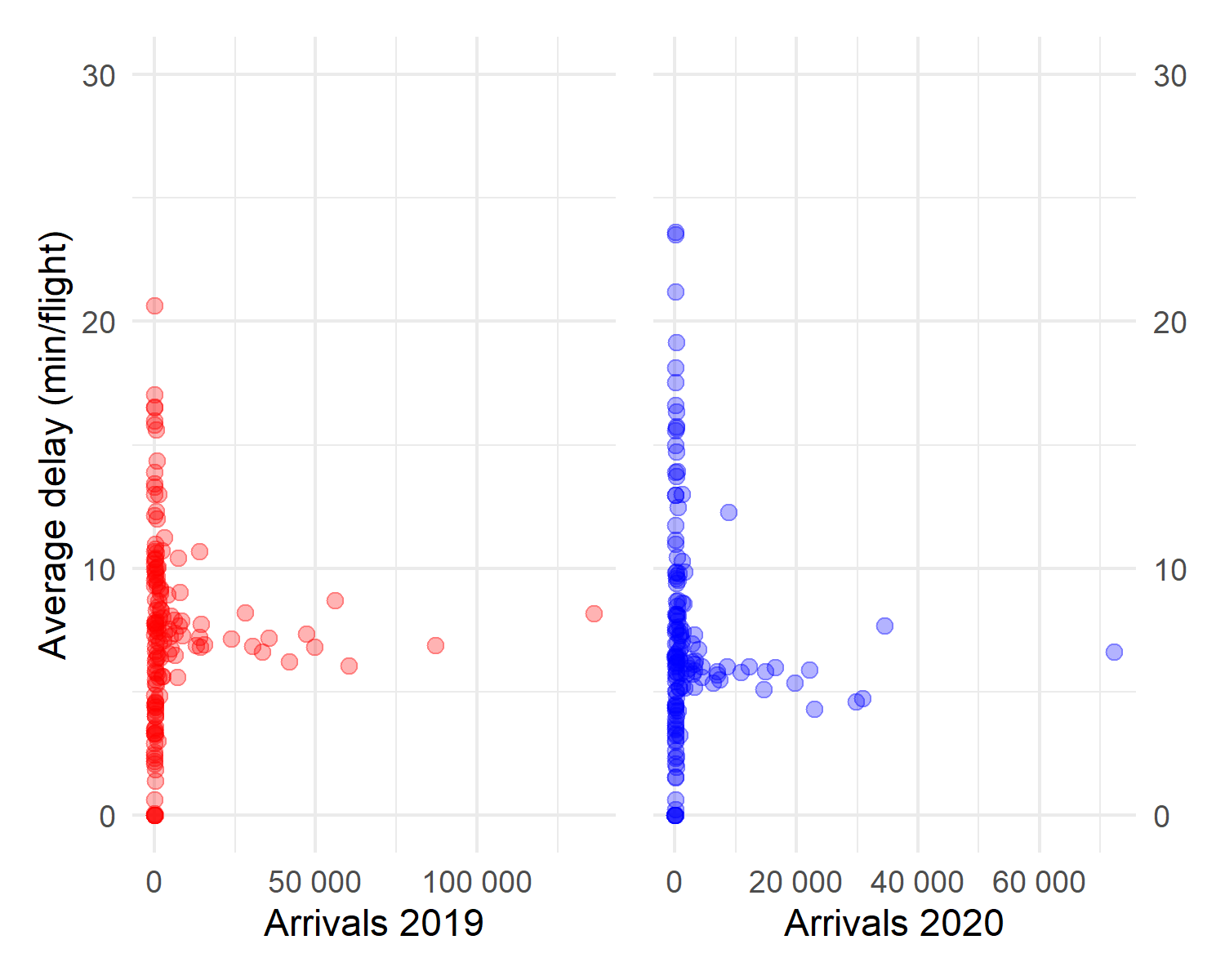
##   
## Paired t-test  
##   
## data: mean\_delays\_top$AVG\_DELAY\_2019 and mean\_delays\_top$AVG\_DELAY\_2020  
## t = 5.5683, df = 49, p-value = 1.075e-06  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.7731393 1.6463164  
## sample estimates:  
## mean of the differences   
## 1.209728

## # A tibble: 100 x 3  
## AIRPORT YEAR AVG\_DELAY  
## <fct> <chr> <dbl>  
## 1 SBGR 2019 8.18  
## 2 SBGR 2020 6.61  
## 3 SBSP 2019 6.90  
## 4 SBSP 2020 4.61  
## 5 SBBR 2019 6.08  
## 6 SBBR 2020 4.75  
## 7 SBKP 2019 8.72  
## 8 SBKP 2020 7.69  
## 9 SBCF 2019 6.82  
## 10 SBCF 2020 5.89  
## # ... with 90 more rows

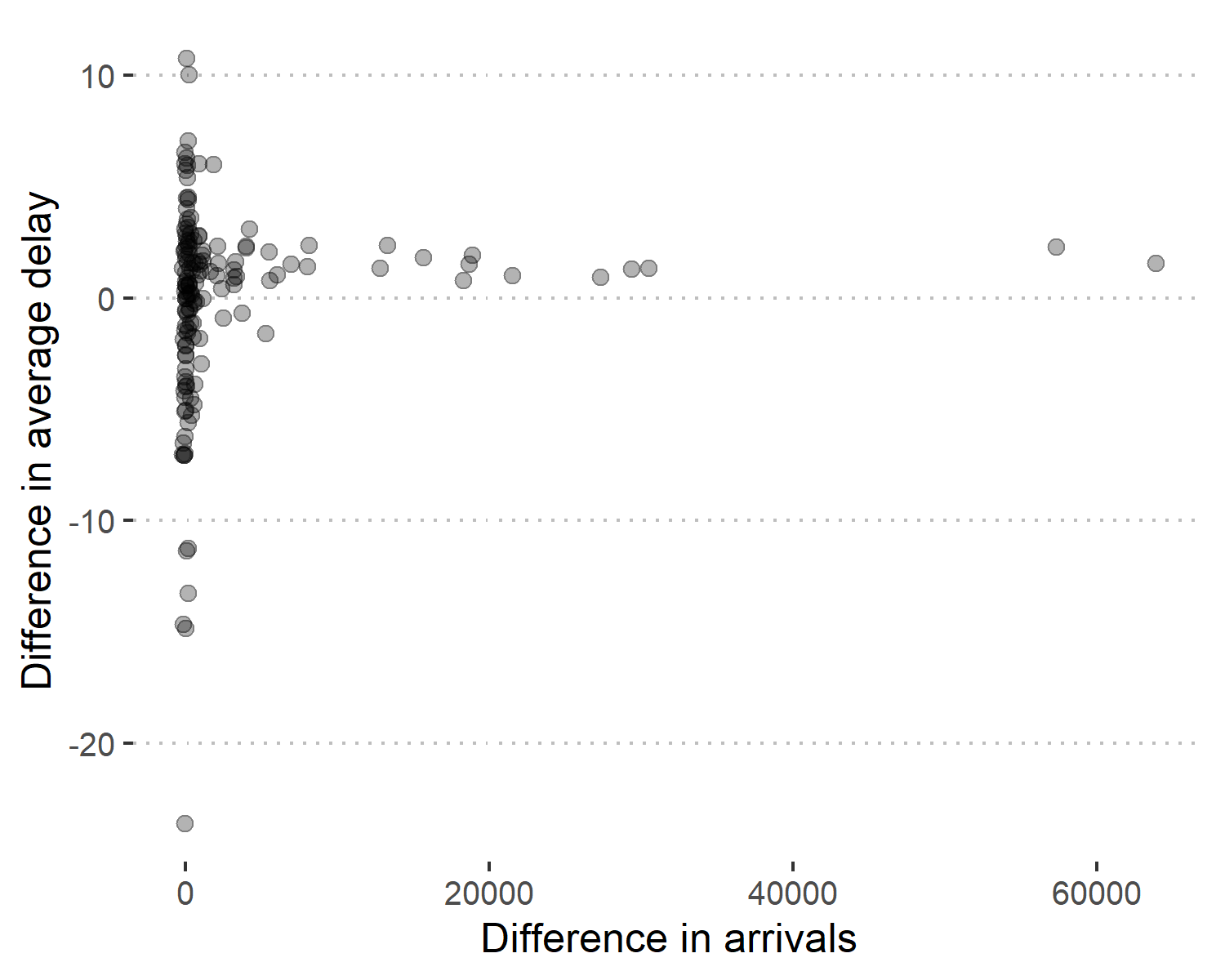
## # A tibble: 2 x 5  
## YEAR count mean median sd  
## <chr> <int> <dbl> <dbl> <dbl>  
## 1 2019 50 7.64 7.35 1.36  
## 2 2020 50 6.43 6.00 1.71



## # A tibble: 150 x 5  
## AIRPORT ARRIVALS\_2019 AVG\_DELAY\_2019 ARRIVALS\_2020 AVG\_DELAY\_2020  
## <fct> <int> <dbl> <int> <dbl>  
## 1 SBGR 136214 8.18 72336 6.61  
## 2 SBSP 87161 6.90 29852 4.61  
## 3 SBBR 60229 6.08 30855 4.75  
## 4 SBKP 55956 8.72 34443 7.69  
## 5 SBCF 49512 6.82 22158 5.89  
## 6 SBGL 47001 7.34 16512 5.98  
## 7 SBRJ 41753 6.24 22884 4.31  
## 8 SBRF 35325 7.19 19670 5.38  
## 9 SBPA 33333 6.64 14633 5.11  
## 10 SBCT 30445 6.85 12152 6.04  
## # ... with 140 more rows



## # A tibble: 150 x 5  
## AIRPORT ARRIVALS\_2019 AVG\_DELAY\_2019 ARRIVALS\_2020 AVG\_DELAY\_2020  
## <fct> <int> <dbl> <int> <dbl>  
## 1 SBGR 136214 8.18 72336 6.61  
## 2 SBSP 87161 6.90 29852 4.61  
## 3 SBBR 60229 6.08 30855 4.75  
## 4 SBKP 55956 8.72 34443 7.69  
## 5 SBCF 49512 6.82 22158 5.89  
## 6 SBGL 47001 7.34 16512 5.98  
## 7 SBRJ 41753 6.24 22884 4.31  
## 8 SBRF 35325 7.19 19670 5.38  
## 9 SBPA 33333 6.64 14633 5.11  
## 10 SBCT 30445 6.85 12152 6.04  
## # ... with 140 more rows



## Average flight time between city pairs correlation test

## # A tibble: 327 x 5  
## CITY\_PAIR FLTS\_2019 AVG\_FLT\_TIME\_2019 FLTS\_2020 AVG\_FLT\_TIME\_2020  
## <chr> <int> <dbl> <int> <dbl>  
## 1 SBSP-SBRJ 17671 0.986 7316 0.979  
## 2 SBRJ-SBSP 17616 1.05 7298 1.05   
## 3 SBPA-SBGR 8416 1.71 4390 1.65   
## 4 SBGR-SBPA 8377 1.73 4406 1.71   
## 5 SBSP-SBBR 7771 1.71 2899 1.69   
## 6 SBBR-SBSP 7758 1.70 2894 1.67   
## 7 SBCF-SBSP 7725 1.30 3017 1.29   
## 8 SBSP-SBCF 7655 1.22 3008 1.20   
## 9 SBSP-SBPA 6917 1.64 2282 1.61   
## 10 SBPA-SBSP 6864 1.54 2260 1.50   
## # ... with 317 more rows

## CITY\_PAIR FLTS\_2019 AVG\_FLT\_TIME\_2019 FLTS\_2020   
## Length:327 Min. : 738 Min. : 0.5562 Min. : 114.0   
## Class :character 1st Qu.: 1155 1st Qu.: 1.2271 1st Qu.: 463.5   
## Mode :character Median : 1533 Median : 1.6142 Median : 756.0   
## Mean : 2184 Mean : 2.2741 Mean :1021.7   
## 3rd Qu.: 2502 3rd Qu.: 2.6130 3rd Qu.:1176.0   
## Max. :17671 Max. :11.6768 Max. :7316.0   
## AVG\_FLT\_TIME\_2020  
## Min. : 0.5453   
## 1st Qu.: 1.2083   
## Median : 1.6058   
## Mean : 2.2608   
## 3rd Qu.: 2.6364   
## Max. :11.6637

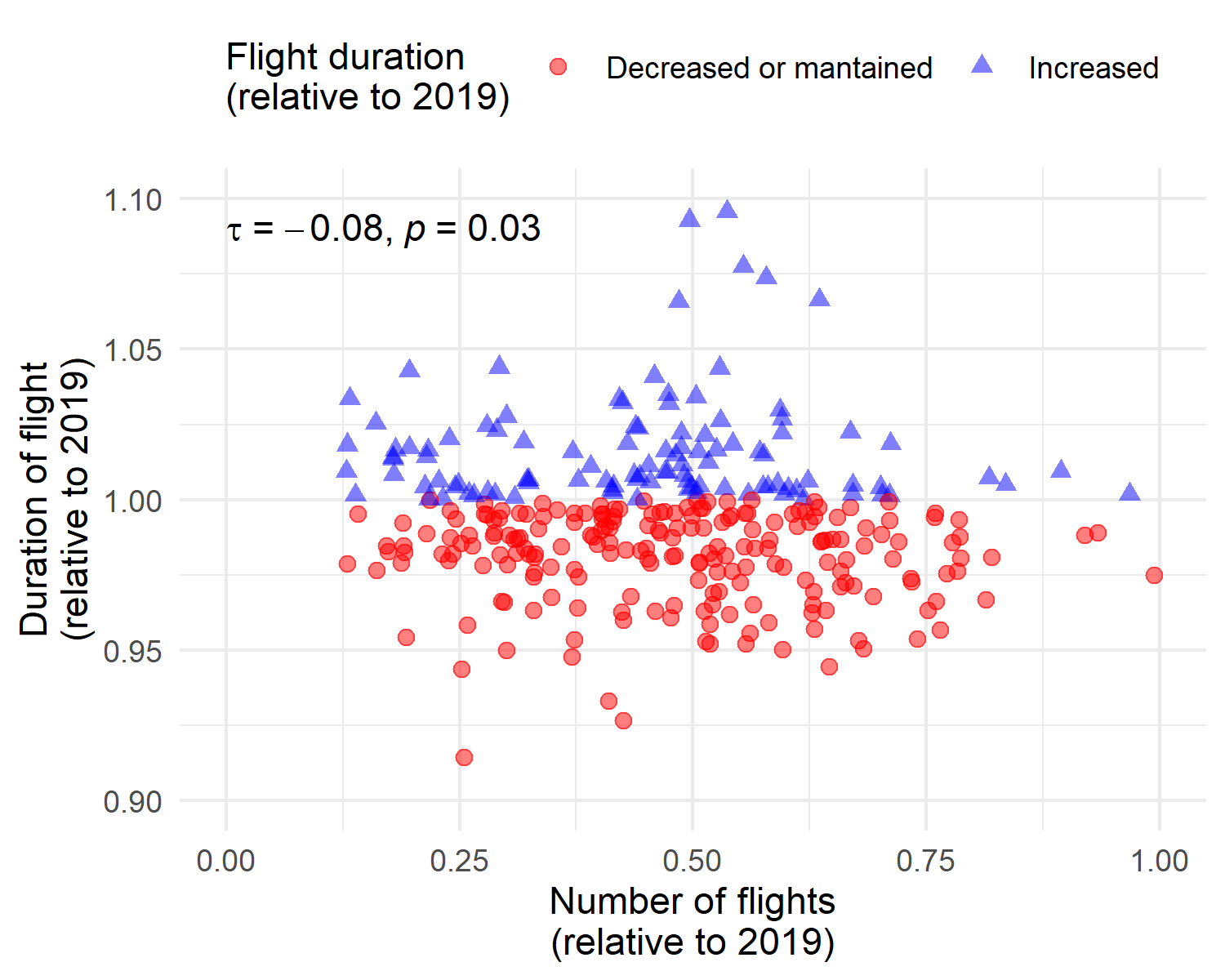
## # A tibble: 327 x 3  
## CITY\_PAIR NORM\_FLT\_2020 NORM\_FLT\_TIME\_2020  
## <chr> <dbl> <dbl>  
## 1 SBSP-SBRJ 0.414 0.993  
## 2 SBRJ-SBSP 0.414 1.00   
## 3 SBPA-SBGR 0.522 0.969  
## 4 SBGR-SBPA 0.526 0.984  
## 5 SBSP-SBBR 0.373 0.992  
## 6 SBBR-SBSP 0.373 0.977  
## 7 SBCF-SBSP 0.391 0.988  
## 8 SBSP-SBCF 0.393 0.988  
## 9 SBSP-SBPA 0.330 0.981  
## 10 SBPA-SBSP 0.329 0.974  
## # ... with 317 more rows

## CITY\_PAIR NORM\_FLT\_2020 NORM\_FLT\_TIME\_2020  
## Length:327 Min. :0.1285 Min. :0.8782   
## Class :character 1st Qu.:0.3241 1st Qu.:0.9804   
## Mode :character Median :0.4775 Median :0.9937   
## Mean :0.4706 Mean :0.9956   
## 3rd Qu.:0.5899 3rd Qu.:1.0046   
## Max. :1.0601 Max. :1.3524

## CITY\_PAIR NORM\_FLT\_2020 NORM\_FLT\_TIME\_2020 decrease   
## Length:327 Min. :0.1285 Min. :0.8782 Mode :logical   
## Class :character 1st Qu.:0.3241 1st Qu.:0.9804 FALSE:1   
## Mode :character Median :0.4775 Median :0.9937 TRUE :326   
## Mean :0.4706 Mean :0.9956   
## 3rd Qu.:0.5899 3rd Qu.:1.0046   
## Max. :1.0601 Max. :1.3524

##   
## Shapiro-Wilk normality test  
##   
## data: norm\_city\_pairs$NORM\_FLT\_2020  
## W = 0.98655, p-value = 0.003889

##   
## Shapiro-Wilk normality test  
##   
## data: norm\_city\_pairs$NORM\_FLT\_TIME\_2020  
## W = 0.72646, p-value < 2.2e-16



##   
## Kendall's rank correlation tau  
##   
## data: ft$NORM\_FLT\_2020 and ft$NORM\_FLT\_TIME\_2020  
## z = -2.1387, p-value = 0.03246  
## alternative hypothesis: true tau is not equal to 0  
## sample estimates:  
## tau   
## -0.0793338