

Data Transformation and Visualization

Computational Statistics

September 25, 2023

```
# Load the required packages
library(tidyverse)
library(nycflights13)
```

Basic data transformation

1a. Get all the flights with "JFK" as the origin airport in the month of September.

Solution:

```
library(dplyr)
nycflights13::flights
```

```
## # A tibble: 336,776 x 19
##   year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##   <int> <int> <int>   <int>         <int>         <dbl>   <int>         <int>
## 1  2013     1     1     517           515           2     830           819
## 2  2013     1     1     533           529           4     850           830
## 3  2013     1     1     542           540           2     923           850
## 4  2013     1     1     544           545          -1    1004          1022
## 5  2013     1     1     554           600          -6     812           837
## 6  2013     1     1     554           558          -4     740           728
## 7  2013     1     1     555           600          -5     913           854
## 8  2013     1     1     557           600          -3     709           723
## 9  2013     1     1     557           600          -3     838           846
## 10 2013     1     1     558           600          -2     753           745
## # i 336,766 more rows
## # i 11 more variables: arr_delay <dbl>, carrier <chr>, flight <int>,
## #   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>,
## #   hour <dbl>, minute <dbl>, time_hour <dtm>
```

```
#Question 1A to filter the flights with origin as JFK and month as September
filter(flights, origin=="JFK", month==9)
```

```
## # A tibble: 8,908 x 19
##   year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##   <int> <int> <int>   <int>         <int>         <dbl>   <int>         <int>
## 1  2013     9     1       9           2359          10     343           340
## 2  2013     9     1     117           2245         152     218           2359
## 3  2013     9     1     537           545          -8     849           855
## 4  2013     9     1     537           545          -8     906           921
## 5  2013     9     1     557           600          -3     706           716
## 6  2013     9     1     602           605          -3     853           855
## 7  2013     9     1     604           605          -1     823           839
## 8  2013     9     1     614           611           3     951           954
## 9  2013     9     1     617           620          -3     829           840
## 10 2013     9     1     621           630          -9     903           924
## # i 8,898 more rows
## # i 11 more variables: arr_delay <dbl>, carrier <chr>, flight <int>,
## #   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>,
## #   hour <dbl>, minute <dbl>, time_hour <dtm>
```

1b. Calculate the average arrival and departure delay for all flights with "JFK" as the origin airport in the month of September.

Solution:

```
#filter the flights with origin as JFK and month=September
question1b <- filter(flights, origin=="JFK", month==9)

#Calculate the average delays using the summarize function
average_delays_sb<-summarize(
  question1b,
  "Average Arrrival Delay"=avg_arr_delay_sb<-mean(arr_delay,na.rm=TRUE),
  "Average Departure Delay"=avg_dep_delay_sb<-mean(dep_delay,na.rm=TRUE)
)
#Call the results
average_delays_sb
```

```
## # A tibble: 1 x 2
##   'Average Arrrival Delay' 'Average Departure Delay'
##               <dbl>               <dbl>
## 1                -4.46                6.64
```

1c. Get the number of trips corresponding to each origin airport ("JFK", "LGA", "EWR").

Solution:

```
#First group the flights by the origin
grouping_by_origin_1c <- group_by(flights, "Origin Airport"=origin)
#find the number of trips against each origin using n() function
summarize(grouping_by_origin_1c, "Number of Flights" = n())
```

```
## # A tibble: 3 x 2
##   'Origin Airport' 'Number of Flights'
##   <chr>           <int>
## 1 EWR             120835
## 2 JFK             111279
## 3 LGA             104662
```

Making sense of data

2a. Select all columns with a name that contains "arr_". Create new variables arr_hour and arr_min for the arrival time in hours and minutes, and sched_arr_hour and sched_arr_min for the scheduled arrival time in hours and minutes. Save the result to a new tibble/data.frame, named arrival. arrival should have the same number of rows as in flights, and 7 columns. **Hint:** Use %/% and %%.

Solution:

```
#Using the pipe operation to pass the results of one expression to the following
arrival <- flights |>
  mutate(
    arr_hour = flights$arr_time %/% 100, #gives the hour of the arr_time
    arr_min = flights$arr_time %% 100, #gives the minutes of the arr_time
    sched_arr_hour = flights$sched_arr_time %/% 100, #gives the minutes of the sched_arr_time
    sched_arr_min = flights$sched_arr_time %% 100 #gives the minutes of the sched_arr_time
  )|>
  select(
    contains("arr_") #Selects only those columns which contain "arr_" in the column name
  )
#display arrival
arrival
```

```
## # A tibble: 336,776 x 7
##   arr_time sched_arr_time arr_delay arr_hour arr_min sched_arr_hour
##   <int>      <int>      <dbl>   <dbl>   <dbl>       <dbl>
## 1     830         819        11       8      30           8
## 2     850         830        20       8      50           8
## 3     923         850        33       9      23           8
## 4    1004        1022       -18      10       4          10
## 5     812         837       -25       8     12           8
```

```
## 6      740      728      12      7      40      7
## 7      913      854      19      9      13      8
## 8      709      723     -14      7      9      7
## 9      838      846      -8      8     38      8
## 10     753      745       8      7     53      7
## # i 336,766 more rows
## # i 1 more variable: sched_arr_min <dbl>
```

```
## The number of rows in the data frame Arrival are : 336776
```

```
## Are the number of rows in the data frame Arrival same as in Flights : TRUE
```

2b. Calculate the arrival delay using `arr_hour`, `arr_min`, `sched_arr_hour` and `sched_arr_min` in `arrival`, and add a new variable `arr_delay2` to store the result. `arr_delay2` should be equal to `arr_delay` in most (if not all) cases. Explain the remaining discrepancy in your result. **Hint:** Use `filter()` to verify your answer. You may also want to use `relocate()` to ease comparison.

Solution:

```
#creating a new data frame with new column arr_delay2
arrival_new <- arrival |>
  mutate(
    arrival_in_min = (arr_hour * 60) + arr_min,
    scheduled_arrival_in_min = (sched_arr_hour * 60) + sched_arr_min,
    arr_delay2 = arrival_in_min - scheduled_arrival_in_min #gives delayed time in minutes
  )|>
  relocate(arr_delay2, .after = arr_delay)

# View the rows of the resulting data frame
arrival_new
```

```
## # A tibble: 336,776 x 10
##   arr_time sched_arr_time arr_delay arr_delay2 arr_hour arr_min sched_arr_hour
##   <int>      <int>      <dbl>      <dbl>      <dbl>  <dbl>      <dbl>
## 1      830        819        11         11         8       30         8
## 2      850        830        20         20         8       50         8
## 3      923        850        33         33         9       23         8
## 4     1004       1022       -18        -18        10        4        10
## 5      812        837       -25        -25         8       12         8
## 6      740        728        12         12         7       40         7
## 7      913        854        19         19         9       13         8
## 8      709        723       -14        -14         7        9         7
## 9      838        846        -8         -8         8       38         8
## 10     753        745         8          8         7       53         7
## # i 336,766 more rows
## # i 3 more variables: sched_arr_min <dbl>, arrival_in_min <dbl>,
## #   scheduled_arrival_in_min <dbl>
```

```
#filter rows where arr_delay is not equal to arr_delay2
differences_2b<-filter(arrival_new, arr_delay!=arr_delay2)
differences_2b
```

```
## # A tibble: 7,295 x 10
##   arr_time sched_arr_time arr_delay arr_delay2 arr_hour arr_min sched_arr_hour
##   <int>      <int>      <dbl>      <dbl>      <dbl>  <dbl>      <dbl>
## 1    1001        1950        851       -589         10      1         19
## 2    2337          5       -28       1412         23     37          0
## 3      8       2359         9     -1431          0      8         23
## 4    2345         17       -32       1408         23     45          0
## 5      6       2323        43     -1397          0      6         23
## 6     20       2352        28     -1412          0     20         23
## 7     46       2040       246     -1194          0     46         20
## 8     32       2357        35     -1405          0     32         23
## 9     21       2110       191     -1249          0     21         21
## 10    22       2313        69     -1371          0     22         23
## # i 7,285 more rows
## # i 3 more variables: sched_arr_min <dbl>, arrival_in_min <dbl>,
## #   scheduled_arrival_in_min <dbl>
```

INFERENCE: From the tibble 'differences_2b' we can infer that 7,295 out of the 336,766 flights have discrepancies in the given arrival delay 'arr_delay' and calculated arrival delay i.e 'arr_delay2'. Negative sign in the 'arr_delay2' column indicates that the flight has arrived early than the scheduled time. Conversely flights have arrived late if the arr_delay2 is a positive number.

1. The values of calculated arr_delay2 will not be appropriate when the arr_time or sched_arr_time is between 0000-0059 (i.e 12am-12:59am) since basic arithmetic operations fail to identify this time as 12 hours. Therefore we see the discrepancy in Rows 2,3,5, etc.

2. At times, the given arr_delay does not match the arr_delay2 as arr_time does not indicate which day the flight has arrived. Therefore, we see that arr_delay2 is always calculated considering the flight has arrived on the same day. For instance, in Row 1 the scheduled arrival time is 1950 hours and the actual arrival time is 1001 hours. The arr_delay indicates a positive delay therefore it is understood that the flight has arrived at 10:01am the next day. But, arr_delay2 indicates a negative delay, that means the flight has arrived early on the same day at 10:01am.

2c. Get flights that arrived in 2014. Explain your reasoning.

Solution : From the dataset `nycflights13::flights`, as there are no records with flight data in 2014, we can extract the flight information by filtering out the flights that arrived on December 31st after 23:59 in the following way.

```
flight2014_2c<-filter(flights, arr_time<dep_time,month==12,day==31)
cat("The flight information is as follows")
```

```
## The flight information is as follows
```

```
t(flight2014_2c)
```

```
##           [,1]           [,2]
## year      "2013"         "2013"
## month     "12"          "12"
## day       "31"          "31"
## dep_time  "1931"         "2012"
## sched_dep_time "1932"     "1900"
## dep_delay  "-1"         "72"
## arr_time   " 28"        " 53"
## sched_arr_time " 28"     "2359"
## arr_delay   " 0"        " 54"
## carrier    "B6"         "AA"
## flight     " 803"        "1029"
## tailnum    "N592JB"      "N5DSAA"
## origin     "JFK"         "JFK"
## dest       "SJU"         "SJU"
## air_time   "193"         "193"
## distance   "1598"        "1598"
## hour       "19"          "19"
## minute     "32"          " 0"
## time_hour  "2013-12-31 19:00:00" "2013-12-31 19:00:00"
##           [,3]           [,4]
## year      "2013"         "2013"
## month     "12"          "12"
## day       "31"          "31"
## dep_time  "2021"         "2039"
## sched_dep_time "2017"     "2046"
## dep_delay   " 4"         "-7"
## arr_time    "101"        " 7"
## sched_arr_time " 116"     "2358"
## arr_delay   "-15"        " 9"
## carrier     "UA"         "B6"
## flight      "1244"        " 329"
## tailnum     "N15710"      "N571JB"
## origin      "EWR"         "JFK"
## dest        "SJU"         "RSW"
## air_time    "204"         "179"
## distance    "1608"        "1074"
## hour        "20"          "20"
## minute      "17"          "46"
## time_hour   "2013-12-31 20:00:00" "2013-12-31 20:00:00"
##           [,5]           [,6]
## year      "2013"         "2013"
## month     "12"          "12"
```

## day	"31"	"31"
## dep_time	"2054"	"2054"
## sched_dep_time	"2101"	"2035"
## dep_delay	"-7"	"19"
## arr_time	" 22"	" 2"
## sched_arr_time	"2346"	"2344"
## arr_delay	" 36"	" 18"
## carrier	"B6"	"DL"
## flight	" 499"	"2065"
## tailnum	"N563JB"	"N391DA"
## origin	"LGA"	"JFK"
## dest	"MCO"	"FLL"
## air_time	"154"	"162"
## distance	" 950"	"1069"
## hour	"21"	"20"
## minute	" 1"	"35"
## time_hour	"2013-12-31 21:00:00"	"2013-12-31 20:00:00"
##	[,7]	[,8]
## year	"2013"	"2013"
## month	"12"	"12"
## day	"31"	"31"
## dep_time	"2059"	"2115"
## sched_dep_time	"2106"	"2050"
## dep_delay	"-7"	"25"
## arr_time	" 10"	"144"
## sched_arr_time	" 7"	" 141"
## arr_delay	" 3"	" 3"
## carrier	"B6"	"UA"
## flight	" 705"	"1071"
## tailnum	"N323JB"	"N76519"
## origin	"EWR"	"EWR"
## dest	"FLL"	"BQN"
## air_time	"160"	"193"
## distance	"1065"	"1585"
## hour	"21"	"20"
## minute	" 6"	"50"
## time_hour	"2013-12-31 21:00:00"	"2013-12-31 20:00:00"
##	[,9]	[,10]
## year	"2013"	"2013"
## month	"12"	"12"
## day	"31"	"31"
## dep_time	"2116"	"2118"
## sched_dep_time	"2045"	"2123"
## dep_delay	"31"	"-5"
## arr_time	" 34"	" 12"

## sched_arr_time	"2359"	"2352"
## arr_delay	" 35"	" 20"
## carrier	"B6"	"B6"
## flight	"1295"	" 97"
## tailnum	"N645JB"	"N584JB"
## origin	"JFK"	"JFK"
## dest	"AUS"	"DEN"
## air_time	"235"	"249"
## distance	"1521"	"1626"
## hour	"20"	"21"
## minute	"45"	"23"
## time_hour	"2013-12-31 20:00:00"	"2013-12-31 21:00:00"
##	[,11]	[,12]
## year	"2013"	"2013"
## month	"12"	"12"
## day	"31"	"31"
## dep_time	"2120"	"2122"
## sched_dep_time	"2129"	"2130"
## dep_delay	"-9"	"-8"
## arr_time	" 11"	" 2"
## sched_arr_time	" 22"	"2359"
## arr_delay	"-11"	" 3"
## carrier	"B6"	"B6"
## flight	"2043"	"1677"
## tailnum	"N556JB"	"N307JB"
## origin	"EWR"	"JFK"
## dest	"PBI"	"JAX"
## air_time	"155"	"138"
## distance	"1023"	" 828"
## hour	"21"	"21"
## minute	"29"	"30"
## time_hour	"2013-12-31 21:00:00"	"2013-12-31 21:00:00"
##	[,13]	[,14]
## year	"2013"	"2013"
## month	"12"	"12"
## day	"31"	"31"
## dep_time	"2128"	"2132"
## sched_dep_time	"2135"	"2130"
## dep_delay	"-7"	" 2"
## arr_time	"101"	" 41"
## sched_arr_time	" 55"	" 18"
## arr_delay	" 6"	" 23"
## carrier	"AA"	"B6"
## flight	" 185"	"1371"
## tailnum	"N332AA"	"N599JB"

## origin	"JFK"	"LGA"
## dest	"LAX"	"FLL"
## air_time	"360"	"160"
## distance	"2475"	"1076"
## hour	"21"	"21"
## minute	"35"	"30"
## time_hour	"2013-12-31 21:00:00"	"2013-12-31 21:00:00"
##	[,15]	[,16]
## year	"2013"	"2013"
## month	"12"	"12"
## day	"31"	"31"
## dep_time	"2134"	"2138"
## sched_dep_time	"2135"	"2141"
## dep_delay	"-1"	"-3"
## arr_time	" 21"	" 50"
## sched_arr_time	" 19"	" 37"
## arr_delay	" 2"	" 13"
## carrier	"B6"	"B6"
## flight	" 611"	" 425"
## tailnum	"N554JB"	"N705JB"
## origin	"JFK"	"JFK"
## dest	"LAS"	"TPA"
## air_time	"320"	"167"
## distance	"2248"	"1005"
## hour	"21"	"21"
## minute	"35"	"41"
## time_hour	"2013-12-31 21:00:00"	"2013-12-31 21:00:00"
##	[,17]	[,18]
## year	"2013"	"2013"
## month	"12"	"12"
## day	"31"	"31"
## dep_time	"2141"	"2155"
## sched_dep_time	"2145"	"2039"
## dep_delay	"-4"	"76"
## arr_time	" 29"	"253"
## sched_arr_time	" 19"	"2355"
## arr_delay	" 10"	NA
## carrier	"B6"	"B6"
## flight	"2227"	"1205"
## tailnum	"N274JB"	"N627JB"
## origin	"EWR"	"JFK"
## dest	"MCO"	"PDX"
## air_time	"144"	NA
## distance	" 937"	"2454"
## hour	"21"	"20"

## minute	"45"	"39"
## time_hour	"2013-12-31 21:00:00"	"2013-12-31 20:00:00"
##	[,19]	[,20]
## year	"2013"	"2013"
## month	"12"	"12"
## day	"31"	"31"
## dep_time	"2155"	"2159"
## sched_dep_time	"2150"	"2155"
## dep_delay	" 5"	" 4"
## arr_time	"110"	" 55"
## sched_arr_time	" 51"	" 46"
## arr_delay	" 19"	" 9"
## carrier	"B6"	"B6"
## flight	"1901"	"2053"
## tailnum	"N729JB"	"N593JB"
## origin	"JFK"	"JFK"
## dest	"FLL"	"PBI"
## air_time	"164"	"155"
## distance	"1069"	"1028"
## hour	"21"	"21"
## minute	"50"	"55"
## time_hour	"2013-12-31 21:00:00"	"2013-12-31 21:00:00"
##	[,21]	[,22]
## year	"2013"	"2013"
## month	"12"	"12"
## day	"31"	"31"
## dep_time	"2206"	"2211"
## sched_dep_time	"2110"	"2159"
## dep_delay	"56"	"12"
## arr_time	" 44"	"100"
## sched_arr_time	"2339"	" 45"
## arr_delay	" 65"	" 15"
## carrier	"B6"	"B6"
## flight	" 775"	"1183"
## tailnum	"N184JB"	"N715JB"
## origin	"JFK"	"JFK"
## dest	"MSY"	"MCO"
## air_time	"195"	"148"
## distance	"1182"	" 944"
## hour	"21"	"21"
## minute	"10"	"59"
## time_hour	"2013-12-31 21:00:00"	"2013-12-31 21:00:00"
##	[,23]	[,24]
## year	"2013"	"2013"
## month	"12"	"12"

## day	"31"	"31"
## dep_time	"2218"	"2310"
## sched_dep_time	"2219"	"2255"
## dep_delay	"-1"	"15"
## arr_time	"315"	" 7"
## sched_arr_time	" 304"	"2356"
## arr_delay	" 11"	" 11"
## carrier	"B6"	"B6"
## flight	"1203"	" 718"
## tailnum	"N625JB"	"N279JB"
## origin	"JFK"	"JFK"
## dest	"SJU"	"BOS"
## air_time	"202"	" 40"
## distance	"1598"	" 187"
## hour	"22"	"22"
## minute	"19"	"55"
## time_hour	"2013-12-31 22:00:00"	"2013-12-31 22:00:00"
##	[,25]	[,26]
## year	"2013"	"2013"
## month	"12"	"12"
## day	"31"	"31"
## dep_time	"2321"	"2328"
## sched_dep_time	"2250"	"2330"
## dep_delay	"31"	"-2"
## arr_time	" 46"	"412"
## sched_arr_time	" 8"	" 409"
## arr_delay	" 38"	" 3"
## carrier	"B6"	"B6"
## flight	"2002"	"1389"
## tailnum	"N179JB"	"N651JB"
## origin	"JFK"	"EWR"
## dest	"BUF"	"SJU"
## air_time	" 66"	"198"
## distance	" 301"	"1608"
## hour	"22"	"23"
## minute	"50"	"30"
## time_hour	"2013-12-31 22:00:00"	"2013-12-31 23:00:00"
##	[,27]	[,28]
## year	"2013"	"2013"
## month	"12"	"12"
## day	"31"	"31"
## dep_time	"2332"	"2355"
## sched_dep_time	"2245"	"2359"
## dep_delay	"47"	"-4"
## arr_time	" 58"	"430"

```

## sched_arr_time " 3" " 440"
## arr_delay " 55" "-10"
## carrier "B6" "B6"
## flight " 486" "1503"
## tailnum "N334JB" "N509JB"
## origin "JFK" "JFK"
## dest "ROC" "SJU"
## air_time " 60" "195"
## distance " 264" "1598"
## hour "22" "23"
## minute "45" "59"
## time_hour "2013-12-31 22:00:00" "2013-12-31 23:00:00"
## [,29]
## year "2013"
## month "12"
## day "31"
## dep_time "2356"
## sched_dep_time "2359"
## dep_delay "-3"
## arr_time "436"
## sched_arr_time " 445"
## arr_delay " -9"
## carrier "B6"
## flight " 745"
## tailnum "N665JB"
## origin "JFK"
## dest "PSE"
## air_time "200"
## distance "1617"
## hour "23"
## minute "59"
## time_hour "2013-12-31 23:00:00"

```

The only flight that arrived in 2014 has departed from NYC at 7:15pm on Dec 31st and arrived at the destination SJU at 12am. This flight arrived 12 minutes prior to the actual arrival time. At 12am on 31st december, the year would technically change to 2014, therefore this explains the result.

Data transformation and visualization

3a. Get the total number of trips for each origin, dest pair for carrier code "AA". Add a new variable route to represent each origin/dest pair (e.g., "LGA to STL") and summarize the result as a bar chart showing the total number of trips (in x-axis) for each route (in y-axis). **Hint:** Use `paste()`.

#Solution#

```

library(nycflights13)
library(dplyr)
library(ggplot2)

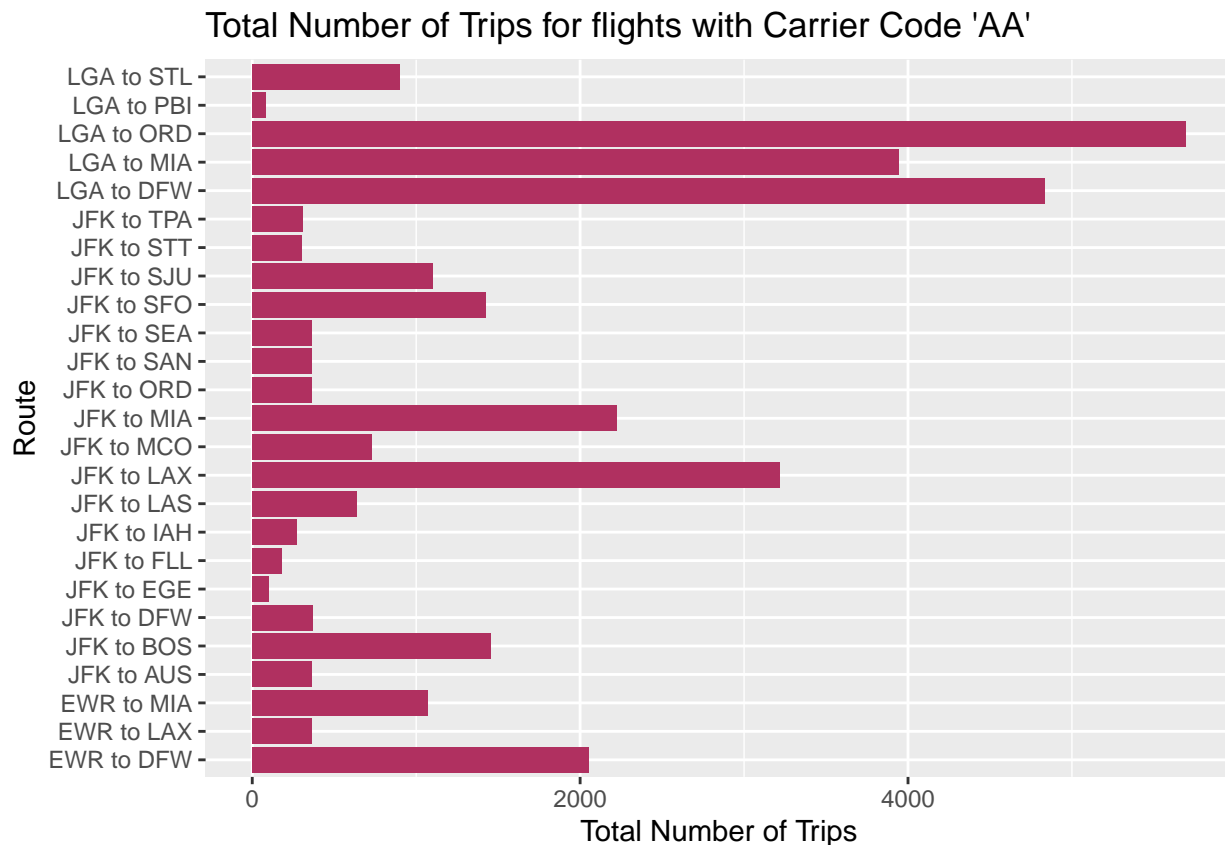
# Firstly filter the flights for carrier code "AA"
carrieraa_flights_3a <- flights |>
  filter(carrier == "AA")

# Assign this to a new variable 'route' by combining 'origin' and 'dest'
carrieraa_flights_3a <- carrieraa_flights_3a|>
  mutate(route = paste(origin, "to", dest))

# Group by 'route' and summarize the total number of trips
flight_summ <- carrieraa_flights_3a |>
  group_by(route) |>
  summarize(total_trips_3a = n())

# Create a bar chart
ggplot(data=flight_summ) +
  geom_bar(stat = "identity", mapping=aes(x = total_trips_3a,y = route), fill="Maroon") +
  labs(y = "Route", x = "Total Number of Trips") +
  ggtitle("Total Number of Trips for flights with Carrier Code 'AA'")

```



3b. Calculate the average speed flown (in MPH) and the average distance flown for each origin, dest pair. Excluding those pairs with average distance over 3,000 miles, in one scatterplot summarize the average speed (in y-axis) versus the average distance (in x-axis). Use different colors to distinguish the three origin airports and add a smooth line with matched color for points corresponding to each origin airport.

Solution

```
# Calculate average speed and average distance for each origin/dest pair
flight_summary <- flights |>
  group_by(origin, dest) |>
  summarize(
    avg_speed = mean(distance / (air_time / 60)),
    avg_distance = mean(distance)
  ) |>
  filter(avg_distance <= 3000) # This removes the pairs with average distance greater than 3,000
```

```
## 'summarise()' has grouped output by 'origin'. You can override using the
## '.groups' argument.
```

```
# Plot with colors and smooth lines
ggplot(flight_summary, mapping=aes(x = avg_distance, y = avg_speed, color = origin)) +
  geom_point() +
  geom_smooth(method = "loess", se=FALSE, linetype = "dotted") +
  labs(
    x = "Average Distance (in miles)",
    y = "Average Speed (MPH)",
    color = "Airport Origin"
  ) +
  scale_color_manual(values = c("JFK" = "purple", "LGA" = "orange", "EWR" = "lightgreen")) +
  ggtitle("Average Speed vs. Average Distance by Origin Airport for NYC Flights 2013")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 207 rows containing non-finite values ('stat_smooth()').
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : span too small. fewer data values than degrees of freedom.
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : pseudoinverse used at 859.86
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : neighborhood radius 104.14
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : reciprocal condition number 0
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : There are other near singularities as well. 8.745e+05
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : Chernobyl! trL>n 5
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : Chernobyl! trL>n 5
```

```
## Warning in sqrt(sum.squares/one.delta): NaNs produced
```

```
## Warning: Removed 207 rows containing missing values ('geom_point()').
```

Average Speed vs. Average Distance by Origin Airport for NYC Flights 201

