Stat 650 Project

Ronald Asseko-Messa

Sunday 13th, October 2019

Loading packages

```
library(pacman)
p_load(tidyverse,knitr, readr, textreadr, forcats, skimr, nycflights13)
```

Loading the data

```
On_Time_Reporting_Carrier1 <- read_csv('/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650 -
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_logical(),
##   Year = col_double(),
##   Quarter = col_double(),
##   Month = col_double(),
##   DayofMonth = col_double(),
##   DayOfWeek = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_character(),
##   DOT_ID_Reporting_Airline = col_double(),
##   IATA_CODE_Reporting_Airline = col_character(),
##   Tail_Number = col_character(),
##   Flight_Number_Reporting_Airline = col_double(),
##   OriginAirportID = col_double(),
##   OriginAirportSeqID = col_double(),
##   OriginCityMarketID = col_double(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
##   OriginWac = col_double()
##   # ... with 44 more columns
## )
```

```
## See spec(...) for full column specifications.
```

```
## Warning: 31561 parsing failures.
##  row                  col           expected actual
## 1861 DivReachedDest       1/0/T/F/TRUE/FALSE 1.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1861 DivActualElapsedTime 1/0/T/F/TRUE/FALSE 284.00 '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1861 DivArrDelay          1/0/T/F/TRUE/FALSE 94.00  '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1861 DivDistance          1/0/T/F/TRUE/FALSE 0.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1861 Div1Airport          1/0/T/F/TRUE/FALSE ABE    '/Users/JaggerCes/Documents/Grad School/Fall 19/
## .... ..................... .................. ...... ........................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier2 <- read_csv('/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650 -
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_character(),
##   IATA_CODE_Reporting_Airline = col_character(),
##   Tail_Number = col_character(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
##   Dest = col_character(),
##   DestCityName = col_character(),
##   DestState = col_character(),
##   DestStateFips = col_character(),
##   DestStateName = col_character(),
##   CRSDepTime = col_character(),
##   DepTime = col_character(),
##   DepTimeBlk = col_character(),
##   WheelsOff = col_character(),
##   WheelsOn = col_character(),
##   CRSArrTime = col_character()
##   # ... with 40 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 124 parsing failures.
##   row              col           expected  actual
## 51172 Div2Airport      1/0/T/F/TRUE/FALSE LAS     '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 51172 Div2AirportID    1/0/T/F/TRUE/FALSE 12889   '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 51172 Div2AirportSeqID 1/0/T/F/TRUE/FALSE 1288903 '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 51172 Div2WheelsOn     1/0/T/F/TRUE/FALSE 2345    '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 51172 Div2TotalGTime   1/0/T/F/TRUE/FALSE 4.00    '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## ..... ................ .................. ....... ........................................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier3 <- read_csv("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650 -
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_character(),
##   IATA_CODE_Reporting_Airline = col_character(),
##   Tail_Number = col_character(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
```

```
##   OriginStateName = col_character(),
##   Dest = col_character(),
##   DestCityName = col_character(),
##   DestState = col_character(),
##   DestStateFips = col_character(),
##   DestStateName = col_character(),
##   CRSDepTime = col_character(),
##   DepTime = col_character(),
##   DepTimeBlk = col_character(),
##   WheelsOff = col_character(),
##   WheelsOn = col_character(),
##   CRSArrTime = col_character()
##   # ... with 40 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 84 parsing failures.
##   row              col           expected  actual
## 30765 Div2Airport      1/0/T/F/TRUE/FALSE PIE     '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 30765 Div2AirportID    1/0/T/F/TRUE/FALSE 14112   '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 30765 Div2AirportSeqID 1/0/T/F/TRUE/FALSE 1411206 '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 30765 Div2WheelsOn     1/0/T/F/TRUE/FALSE 2031    '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 30765 Div2TotalGTime   1/0/T/F/TRUE/FALSE 4.00    '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## ..... ................ .................. ....... ........................................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier4 <- read_csv("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650 -
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_character(),
##   IATA_CODE_Reporting_Airline = col_character(),
##   Tail_Number = col_character(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
##   Dest = col_character(),
##   DestCityName = col_character(),
##   DestState = col_character(),
##   DestStateFips = col_character(),
##   DestStateName = col_character(),
##   CRSDepTime = col_character(),
##   DepTime = col_character(),
##   DepTimeBlk = col_character(),
##   WheelsOff = col_character(),
##   WheelsOn = col_character(),
##   CRSArrTime = col_character()
##   # ... with 44 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 6744 parsing failures.
##  row                  col           expected actual
## 1391 DivReachedDest       1/0/T/F/TRUE/FALSE 1.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/C
## 1391 DivActualElapsedTime 1/0/T/F/TRUE/FALSE 333.00 '/Users/JaggerCes/Documents/Grad School/Fall 19/C
## 1391 DivArrDelay          1/0/T/F/TRUE/FALSE 253.00 '/Users/JaggerCes/Documents/Grad School/Fall 19/C
## 1391 DivDistance          1/0/T/F/TRUE/FALSE 0.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/C
## 1391 Div1WheelsOff        1/0/T/F/TRUE/FALSE 2243   '/Users/JaggerCes/Documents/Grad School/Fall 19/C
## .... ..................... .................. ...... ........................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier5 <- read_csv("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650 -
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_character(),
##   IATA_CODE_Reporting_Airline = col_character(),
##   Tail_Number = col_character(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
##   Dest = col_character(),
##   DestCityName = col_character(),
##   DestState = col_character(),
##   DestStateFips = col_character(),
##   DestStateName = col_character(),
##   CRSDepTime = col_character(),
##   DepTime = col_character(),
##   DepTimeBlk = col_character(),
##   WheelsOff = col_character(),
##   WheelsOn = col_character(),
##   CRSArrTime = col_character()
##   # ... with 38 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 78 parsing failures.
##   row              col           expected  actual
## 29552 Div2Airport      1/0/T/F/TRUE/FALSE LAS     '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 29552 Div2AirportID    1/0/T/F/TRUE/FALSE 12889   '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 29552 Div2AirportSeqID 1/0/T/F/TRUE/FALSE 1288903 '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 29552 Div2WheelsOn     1/0/T/F/TRUE/FALSE 1629    '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 29552 Div2TotalGTime   1/0/T/F/TRUE/FALSE 10.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## ..... ................ .................. ....... ........................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier6 <- read_csv("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650 -
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
```

```
##   .default = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_number(),
##   IATA_CODE_Reporting_Airline = col_number(),
##   Tail_Number = col_character(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
##   Dest = col_character(),
##   DestCityName = col_character(),
##   DestState = col_character(),
##   DestStateFips = col_character(),
##   DestStateName = col_character(),
##   CRSDepTime = col_character(),
##   DepTime = col_character(),
##   DepTimeBlk = col_character(),
##   WheelsOff = col_character(),
##   WheelsOn = col_character(),
##   CRSArrTime = col_character()
##   # ... with 40 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 1116886 parsing failures.
##  row              col           expected  actual
## 8422 Div2Airport      1/0/T/F/TRUE/FALSE MSP     '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 8422 Div2AirportID    1/0/T/F/TRUE/FALSE 13487   '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 8422 Div2AirportSeqID 1/0/T/F/TRUE/FALSE 1348702 '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 8422 Div2WheelsOn     1/0/T/F/TRUE/FALSE 0341    '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 8422 Div2TotalGTime   1/0/T/F/TRUE/FALSE 34.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## .... ............... .................. ....... ..........................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier7 <- read_csv("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650 -
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_number(),
##   IATA_CODE_Reporting_Airline = col_number(),
##   Tail_Number = col_character(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
##   Dest = col_character(),
##   DestCityName = col_character(),
##   DestState = col_character(),
##   DestStateFips = col_character(),
##   DestStateName = col_character(),
```

```
##   CRSDepTime = col_character(),
##   DepTime = col_character(),
##   DepTimeBlk = col_character(),
##   WheelsOff = col_character(),
##   WheelsOn = col_character(),
##   CRSArrTime = col_character()
##   # ... with 41 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 1152322 parsing failures.
##   row              col           expected  actual
## 26995 Div2Airport      1/0/T/F/TRUE/FALSE DEN     '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 26995 Div2AirportID    1/0/T/F/TRUE/FALSE 11292   '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 26995 Div2AirportSeqID 1/0/T/F/TRUE/FALSE 1129202 '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 26995 Div2WheelsOn     1/0/T/F/TRUE/FALSE 0318    '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## 26995 Div2TotalGTime   1/0/T/F/TRUE/FALSE 61.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/CSU
## ..... ................ .................. ....... ..........................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier8 <- read_csv("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650 -
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_character(),
##   IATA_CODE_Reporting_Airline = col_character(),
##   Tail_Number = col_character(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
##   Dest = col_character(),
##   DestCityName = col_character(),
##   DestState = col_character(),
##   DestStateFips = col_character(),
##   DestStateName = col_character(),
##   CRSDepTime = col_character(),
##   DepTime = col_character(),
##   DepTimeBlk = col_character(),
##   WheelsOff = col_character(),
##   WheelsOn = col_character(),
##   CRSArrTime = col_character()
##   # ... with 40 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 12423 parsing failures.
##  row              col           expected actual
## 1550 CancellationCode 1/0/T/F/TRUE/FALSE      A '/Users/JaggerCes/Documents/Grad School/Fall 19/CSUE
## 1866 CancellationCode 1/0/T/F/TRUE/FALSE      B '/Users/JaggerCes/Documents/Grad School/Fall 19/CSUE
## 3852 CancellationCode 1/0/T/F/TRUE/FALSE      B '/Users/JaggerCes/Documents/Grad School/Fall 19/CSUE
```

```
## 4112 CancellationCode 1/0/T/F/TRUE/FALSE      A '/Users/JaggerCes/Documents/Grad School/Fall 19/CSUE
## 4378 CancellationCode 1/0/T/F/TRUE/FALSE      C '/Users/JaggerCes/Documents/Grad School/Fall 19/CSUE
## .... ............... .................. ...... ..........................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier9 <- read_csv("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_logical(),
##   Year = col_double(),
##   Quarter = col_double(),
##   Month = col_double(),
##   DayofMonth = col_double(),
##   DayOfWeek = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_number(),
##   DOT_ID_Reporting_Airline = col_double(),
##   IATA_CODE_Reporting_Airline = col_number(),
##   Tail_Number = col_character(),
##   Flight_Number_Reporting_Airline = col_double(),
##   OriginAirportID = col_double(),
##   OriginAirportSeqID = col_double(),
##   OriginCityMarketID = col_double(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
##   OriginWac = col_double()
##   # ... with 45 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 1067242 parsing failures.
##  row                  col           expected actual
## 1099 DivReachedDest       1/0/T/F/TRUE/FALSE 1.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1099 DivActualElapsedTime 1/0/T/F/TRUE/FALSE 198.00 '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1099 DivArrDelay          1/0/T/F/TRUE/FALSE 78.00  '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1099 DivDistance          1/0/T/F/TRUE/FALSE 0.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1099 Div1Airport          1/0/T/F/TRUE/FALSE FNT    '/Users/JaggerCes/Documents/Grad School/Fall 19/
## .... ..................... .................. ...... ..........................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier10 <- read_csv("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_logical(),
##   Year = col_double(),
##   Quarter = col_double(),
##   Month = col_double(),
##   DayofMonth = col_double(),
```

```
##   DayOfWeek = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_character(),
##   DOT_ID_Reporting_Airline = col_double(),
##   IATA_CODE_Reporting_Airline = col_character(),
##   Tail_Number = col_character(),
##   Flight_Number_Reporting_Airline = col_double(),
##   OriginAirportID = col_double(),
##   OriginAirportSeqID = col_double(),
##   OriginCityMarketID = col_double(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
##   OriginWac = col_double()
##   # ... with 45 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 15266 parsing failures.
##  row                  col           expected actual
## 1447 DivReachedDest       1/0/T/F/TRUE/FALSE 1.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/C
## 1447 DivActualElapsedTime 1/0/T/F/TRUE/FALSE 220.00 '/Users/JaggerCes/Documents/Grad School/Fall 19/C
## 1447 DivArrDelay          1/0/T/F/TRUE/FALSE 229.00 '/Users/JaggerCes/Documents/Grad School/Fall 19/C
## 1447 DivDistance          1/0/T/F/TRUE/FALSE 0.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/C
## 1447 Div1Airport          1/0/T/F/TRUE/FALSE MYR    '/Users/JaggerCes/Documents/Grad School/Fall 19/C
## .... ..................... ................... ...... ........................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier11 <- read_csv("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_logical(),
##   Year = col_double(),
##   Quarter = col_double(),
##   Month = col_double(),
##   DayofMonth = col_double(),
##   DayOfWeek = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_number(),
##   DOT_ID_Reporting_Airline = col_double(),
##   IATA_CODE_Reporting_Airline = col_number(),
##   Tail_Number = col_character(),
##   Flight_Number_Reporting_Airline = col_double(),
##   OriginAirportID = col_double(),
##   OriginAirportSeqID = col_double(),
##   OriginCityMarketID = col_double(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
```

```
##   OriginWac = col_double()
##   # ... with 45 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 1062713 parsing failures.
##  row                  col           expected actual
## 1149 DivReachedDest       1/0/T/F/TRUE/FALSE 1.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1149 DivActualElapsedTime 1/0/T/F/TRUE/FALSE 225.00 '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1149 DivArrDelay          1/0/T/F/TRUE/FALSE 114.00 '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1149 DivDistance          1/0/T/F/TRUE/FALSE 0.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/
## 1149 Div1Airport          1/0/T/F/TRUE/FALSE MCI    '/Users/JaggerCes/Documents/Grad School/Fall 19/
## .... ..................... .................. ...... ..........................................
## See problems(...) for more details.
```

```
On_Time_Reporting_Carrier12 <- read_csv("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650
```

```
## Warning: Missing column names filled in: 'X110' [110]
```

```
## Parsed with column specification:
## cols(
##   .default = col_logical(),
##   Year = col_double(),
##   Quarter = col_double(),
##   Month = col_double(),
##   DayofMonth = col_double(),
##   DayOfWeek = col_double(),
##   FlightDate = col_date(format = ""),
##   Reporting_Airline = col_character(),
##   DOT_ID_Reporting_Airline = col_double(),
##   IATA_CODE_Reporting_Airline = col_character(),
##   Tail_Number = col_character(),
##   Flight_Number_Reporting_Airline = col_double(),
##   OriginAirportID = col_double(),
##   OriginAirportSeqID = col_double(),
##   OriginCityMarketID = col_double(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateFips = col_character(),
##   OriginStateName = col_character(),
##   OriginWac = col_double()
##   # ... with 44 more columns
## )
## See spec(...) for full column specifications.
```

```
## Warning: 22437 parsing failures.
##  row               col           expected actual
## 1108 CancellationCode 1/0/T/F/TRUE/FALSE B      '/Users/JaggerCes/Documents/Grad School/Fall 19/CSUE
## 1110 DivReachedDest   1/0/T/F/TRUE/FALSE 0.00   '/Users/JaggerCes/Documents/Grad School/Fall 19/CSUE
## 1110 DivDistance      1/0/T/F/TRUE/FALSE 239.00 '/Users/JaggerCes/Documents/Grad School/Fall 19/CSUE
## 1110 Div1Airport      1/0/T/F/TRUE/FALSE HOU    '/Users/JaggerCes/Documents/Grad School/Fall 19/CSUE
## 1110 Div1AirportID    1/0/T/F/TRUE/FALSE 12191  '/Users/JaggerCes/Documents/Grad School/Fall 19/CSUE
## .... ................ .................. ...... ..........................................
## See problems(...) for more details.
```

Checking what the head of the data looks like

```
head(On_Time_Reporting_Carrier1)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       1     1         27         6 2018-01-27 UA
## 2  2018       1     1         27         6 2018-01-27 UA
## 3  2018       1     1         27         6 2018-01-27 UA
## 4  2018       1     1         27         6 2018-01-27 UA
## 5  2018       1     1         27         6 2018-01-27 UA
## 6  2018       1     1         27         6 2018-01-27 UA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <lgl>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <lgl>, Div1AirportID <lgl>, Div1AirportSeqID <lgl>,
## #   Div1WheelsOn <lgl>, Div1TotalGTime <lgl>, Div1LongestGTime <lgl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier2)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       1     2          1         4 2018-02-01 F9
## 2  2018       1     2          1         4 2018-02-01 F9
## 3  2018       1     2          1         4 2018-02-01 F9
```

```
## 4  2018       1     2          1         4 2018-02-01 F9
## 5  2018       1     2          1         4 2018-02-01 F9
## 6  2018       1     2          1         4 2018-02-01 F9
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <dbl>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <chr>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier3)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       1     3          1         4 2018-03-01 F9
## 2  2018       1     3          1         4 2018-03-01 F9
## 3  2018       1     3          1         4 2018-03-01 F9
## 4  2018       1     3          1         4 2018-03-01 F9
## 5  2018       1     3          1         4 2018-03-01 F9
## 6  2018       1     3          1         4 2018-03-01 F9
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
```

```
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <dbl>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <chr>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <chr>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier4)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       2     4         26         4 2018-04-26 UA
## 2  2018       2     4         26         4 2018-04-26 UA
## 3  2018       2     4         26         4 2018-04-26 UA
## 4  2018       2     4         26         4 2018-04-26 UA
## 5  2018       2     4         26         4 2018-04-26 UA
## 6  2018       2     4         26         4 2018-04-26 UA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
```

```
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <dbl>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier5)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       2     5          1         2 2018-05-01 F9
## 2  2018       2     5          1         2 2018-05-01 F9
## 3  2018       2     5          1         2 2018-05-01 F9
## 4  2018       2     5          1         2 2018-05-01 F9
## 5  2018       2     5          1         2 2018-05-01 F9
## 6  2018       2     5          1         2 2018-05-01 F9
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <dbl>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <dbl>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <dbl>, Div1TailNum <chr>, Div2Airport <lgl>,
```

```
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier6)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>                <dbl>
## 1  2018       2     6          1         5 2018-06-01                9
## 2  2018       2     6          3         7 2018-06-03                9
## 3  2018       2     6          4         1 2018-06-04                9
## 4  2018       2     6          5         2 2018-06-05                9
## 5  2018       2     6          6         3 2018-06-06                9
## 6  2018       2     6          7         4 2018-06-07                9
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <dbl>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <chr>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <dbl>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
```

```
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier7)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>                <dbl>
## 1  2018       3     7          1         7 2018-07-01                9
## 2  2018       3     7          2         1 2018-07-02                9
## 3  2018       3     7          3         2 2018-07-03                9
## 4  2018       3     7          4         3 2018-07-04                9
## 5  2018       3     7          5         4 2018-07-05                9
## 6  2018       3     7          6         5 2018-07-06                9
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <dbl>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <chr>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <chr>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier8)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       3     8         31         5 2018-08-31 UA
## 2  2018       3     8         31         5 2018-08-31 UA
```

```
## 3  2018       3     8         31         5 2018-08-31 UA
## 4  2018       3     8         31         5 2018-08-31 UA
## 5  2018       3     8         31         5 2018-08-31 UA
## 6  2018       3     8         31         5 2018-08-31 UA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <lgl>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <dbl>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <chr>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <chr>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier9)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>                <dbl>
## 1  2018       3     9          3         1 2018-09-03                9
## 2  2018       3     9          9         7 2018-09-09                9
## 3  2018       3     9         10         1 2018-09-10                9
## 4  2018       3     9         13         4 2018-09-13                9
## 5  2018       3     9         14         5 2018-09-14                9
## 6  2018       3     9         16         7 2018-09-16                9
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <dbl>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
```

```
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <lgl>, Div1AirportID <lgl>, Div1AirportSeqID <lgl>,
## #   Div1WheelsOn <lgl>, Div1TotalGTime <lgl>, Div1LongestGTime <lgl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier10)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       4    10         14         7 2018-10-14 AA
## 2  2018       4    10         15         1 2018-10-15 AA
## 3  2018       4    10         16         2 2018-10-16 AA
## 4  2018       4    10         17         3 2018-10-17 AA
## 5  2018       4    10         18         4 2018-10-18 AA
## 6  2018       4    10         19         5 2018-10-19 AA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
```

```
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <lgl>, Div1AirportID <lgl>, Div1AirportSeqID <lgl>,
## #   Div1WheelsOn <lgl>, Div1TotalGTime <lgl>, Div1LongestGTime <lgl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier11)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>                <dbl>
## 1  2018       4    11         30         5 2018-11-30                9
## 2  2018       4    11         27         2 2018-11-27                9
## 3  2018       4    11         28         3 2018-11-28                9
## 4  2018       4    11         29         4 2018-11-29                9
## 5  2018       4    11         30         5 2018-11-30                9
## 6  2018       4    11          1         4 2018-11-01                9
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <dbl>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <lgl>, Div1AirportID <lgl>, Div1AirportSeqID <lgl>,
## #   Div1WheelsOn <lgl>, Div1TotalGTime <lgl>, Div1LongestGTime <lgl>,
```

```
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
head(On_Time_Reporting_Carrier12)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       4    12         25         2 2018-12-25 WN
## 2  2018       4    12         25         2 2018-12-25 WN
## 3  2018       4    12         25         2 2018-12-25 WN
## 4  2018       4    12         25         2 2018-12-25 WN
## 5  2018       4    12         25         2 2018-12-25 WN
## 6  2018       4    12         25         2 2018-12-25 WN
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <lgl>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <lgl>, Div1AirportID <lgl>, Div1AirportSeqID <lgl>,
## #   Div1WheelsOn <lgl>, Div1TotalGTime <lgl>, Div1LongestGTime <lgl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
```

```
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

Checking what the tail of the data looks like

```
tail(On_Time_Reporting_Carrier1)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       1     1         11         4 2018-01-11 DL
## 2  2018       1     1         11         4 2018-01-11 DL
## 3  2018       1     1         11         4 2018-01-11 DL
## 4  2018       1     1         11         4 2018-01-11 DL
## 5  2018       1     1         11         4 2018-01-11 DL
## 6  2018       1     1         11         4 2018-01-11 DL
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <lgl>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <lgl>, Div1AirportID <lgl>, Div1AirportSeqID <lgl>,
## #   Div1WheelsOn <lgl>, Div1TotalGTime <lgl>, Div1LongestGTime <lgl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier2)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
```

```
## 1  2018       1     2         22         4 2018-02-22 9E
## 2  2018       1     2         23         5 2018-02-23 9E
## 3  2018       1     2         25         7 2018-02-25 9E
## 4  2018       1     2         26         1 2018-02-26 9E
## 5  2018       1     2         27         2 2018-02-27 9E
## 6  2018       1     2         28         3 2018-02-28 9E
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <dbl>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <chr>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier3)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       1     3          1         4 2018-03-01 UA
## 2  2018       1     3          1         4 2018-03-01 UA
## 3  2018       1     3          1         4 2018-03-01 UA
## 4  2018       1     3          1         4 2018-03-01 UA
## 5  2018       1     3          1         4 2018-03-01 UA
## 6  2018       1     3          1         4 2018-03-01 UA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
```

```
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <dbl>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <chr>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <chr>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier4)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       2     4         14         6 2018-04-14 UA
## 2  2018       2     4         14         6 2018-04-14 UA
## 3  2018       2     4         14         6 2018-04-14 UA
## 4  2018       2     4         14         6 2018-04-14 UA
## 5  2018       2     4         14         6 2018-04-14 UA
## 6  2018       2     4         14         6 2018-04-14 UA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
```

```
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <dbl>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier5)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       2     5          8         2 2018-05-08 UA
## 2  2018       2     5          8         2 2018-05-08 UA
## 3  2018       2     5          8         2 2018-05-08 UA
## 4  2018       2     5          8         2 2018-05-08 UA
## 5  2018       2     5          8         2 2018-05-08 UA
## 6  2018       2     5          8         2 2018-05-08 UA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <dbl>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
```

```
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <dbl>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <dbl>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier6)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>                <dbl>
## 1  2018       2     6         21         4 2018-06-21               NA
## 2  2018       2     6         21         4 2018-06-21               NA
## 3  2018       2     6         21         4 2018-06-21               NA
## 4  2018       2     6         21         4 2018-06-21               NA
## 5  2018       2     6         21         4 2018-06-21               NA
## 6  2018       2     6         21         4 2018-06-21               NA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <dbl>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <chr>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <dbl>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
```

```
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier7)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>                <dbl>
## 1  2018       3     7         25         3 2018-07-25               NA
## 2  2018       3     7         25         3 2018-07-25               NA
## 3  2018       3     7         25         3 2018-07-25               NA
## 4  2018       3     7         25         3 2018-07-25               NA
## 5  2018       3     7         25         3 2018-07-25               NA
## 6  2018       3     7         25         3 2018-07-25               NA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <dbl>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <chr>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <chr>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier8)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
```

```
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       3     8         31         5 2018-08-31 AS
## 2  2018       3     8         31         5 2018-08-31 AS
## 3  2018       3     8         31         5 2018-08-31 AS
## 4  2018       3     8         31         5 2018-08-31 AS
## 5  2018       3     8         31         5 2018-08-31 AS
## 6  2018       3     8         31         5 2018-08-31 AS
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <lgl>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <dbl>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <chr>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <chr>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier9)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>                <dbl>
## 1  2018       3     9         11         2 2018-09-11               NA
## 2  2018       3     9         11         2 2018-09-11               NA
## 3  2018       3     9         11         2 2018-09-11               NA
## 4  2018       3     9         11         2 2018-09-11               NA
## 5  2018       3     9         11         2 2018-09-11               NA
## 6  2018       3     9         11         2 2018-09-11               NA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <dbl>, Tail_Number <chr>,
```

```
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <lgl>, Div1AirportID <lgl>, Div1AirportSeqID <lgl>,
## #   Div1WheelsOn <lgl>, Div1TotalGTime <lgl>, Div1LongestGTime <lgl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier10)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       4    10         25         4 2018-10-25 OH
## 2  2018       4    10         26         5 2018-10-26 OH
## 3  2018       4    10         27         6 2018-10-27 OH
## 4  2018       4    10         28         7 2018-10-28 OH
## 5  2018       4    10         29         1 2018-10-29 OH
## 6  2018       4    10         30         2 2018-10-30 OH
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
```

```
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <lgl>, Div1AirportID <lgl>, Div1AirportSeqID <lgl>,
## #   Div1WheelsOn <lgl>, Div1TotalGTime <lgl>, Div1LongestGTime <lgl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier11)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>                <dbl>
## 1  2018       4    11          7         3 2018-11-07               NA
## 2  2018       4    11          7         3 2018-11-07               NA
## 3  2018       4    11          7         3 2018-11-07               NA
## 4  2018       4    11          7         3 2018-11-07               NA
## 5  2018       4    11          7         3 2018-11-07               NA
## 6  2018       4    11          7         3 2018-11-07               NA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <dbl>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
```

```
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <lgl>, Div1AirportID <lgl>, Div1AirportSeqID <lgl>,
## #   Div1WheelsOn <lgl>, Div1TotalGTime <lgl>, Div1LongestGTime <lgl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
tail(On_Time_Reporting_Carrier12)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       4    12         27         4 2018-12-27 DL
## 2  2018       4    12         27         4 2018-12-27 DL
## 3  2018       4    12         27         4 2018-12-27 DL
## 4  2018       4    12         27         4 2018-12-27 DL
## 5  2018       4    12         27         4 2018-12-27 DL
## 6  2018       4    12         27         4 2018-12-27 DL
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <lgl>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <chr>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <lgl>,
## #   DivActualElapsedTime <lgl>, DivArrDelay <lgl>, DivDistance <lgl>,
## #   Div1Airport <lgl>, Div1AirportID <lgl>, Div1AirportSeqID <lgl>,
## #   Div1WheelsOn <lgl>, Div1TotalGTime <lgl>, Div1LongestGTime <lgl>,
## #   Div1WheelsOff <lgl>, Div1TailNum <lgl>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
```

```
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

Changing type of FirstDepTime, Div1WheelsOn, and Reporting_Airline

```
On_Time_Reporting_Carrier1 <- On_Time_Reporting_Carrier1 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier2 <- On_Time_Reporting_Carrier2 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier3 <- On_Time_Reporting_Carrier3 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier4 <- On_Time_Reporting_Carrier4 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier5 <- On_Time_Reporting_Carrier5 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier6 <- On_Time_Reporting_Carrier6 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier7 <- On_Time_Reporting_Carrier7 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
```

```
        Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier8 <- On_Time_Reporting_Carrier8 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier9 <- On_Time_Reporting_Carrier9 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier10 <- On_Time_Reporting_Carrier10 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier11 <- On_Time_Reporting_Carrier11 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))

On_Time_Reporting_Carrier12 <- On_Time_Reporting_Carrier12 %>%
  mutate(FirstDepTime = as.double(FirstDepTime),
         Div1WheelsOn = as.double(Div1WheelsOn),
         Reporting_Airline = as.character(Reporting_Airline),
         IATA_CODE_Reporting_Airline = as.character(IATA_CODE_Reporting_Airline),
         Div1WheelsOff = as.double(Div1WheelsOff))
```

2. Merging the 12 files

```
my_flights <- bind_rows(On_Time_Reporting_Carrier1, On_Time_Reporting_Carrier2, On_Time_Reporting_Carri

head(my_flights)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       1     1         27         6 2018-01-27 UA
## 2  2018       1     1         27         6 2018-01-27 UA
## 3  2018       1     1         27         6 2018-01-27 UA
## 4  2018       1     1         27         6 2018-01-27 UA
## 5  2018       1     1         27         6 2018-01-27 UA
## 6  2018       1     1         27         6 2018-01-27 UA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
```

```
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <dbl>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <dbl>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <dbl>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

There were 22,606 flights that flew from the Bay Area in January 2018. 4194 from OAK, 14135 from SFO and 4277 from SJC

```
bay_flights1 <- my_flights %>%
  filter(Month == 1, Origin == "SFO" | Origin == "OAK" | Origin == "SJC")

bay_flights1 %>% count()
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1 22606
```

```
bay_flights1 %>%
  group_by(Year, Month, Origin) %>%
  count()
```

```
## # A tibble: 3 x 4
## # Groups:   Year, Month, Origin [3]
##    Year Month Origin     n
##   <dbl> <dbl> <chr>  <int>
## 1  2018     1 OAK     4194
## 2  2018     1 SFO    14135
```

```
## 3  2018     1 SJC     4277
```

```
??nycflights13::flights

nycflights13_names <- tribble(
~columns, ~description,
'year', 'Date of departure. - year',
'month', 'Date of departure. - month',
'day', 'Date of departure.',
'dep_time', 'Actual departure and arrival times (format HHMM or HMM), local tz.',
'arr_time', 'Actual departure and arrival times (format HHMM or HMM), local tz.',
'sched_dep_time', 'Scheduled departure and arrival times (format HHMM or HMM), local tz.',
'sched_arr_time', 'Scheduled departure and arrival times (format HHMM or HMM), local tz.',
'dep_delay', 'Departure and arrival delays, in minutes. Negative times represent early departures/arriva
'arr_delay', 'Departure and arrival delays, in minutes. Negative times represent early departures/arriva
'carrier', 'Two letter carrier abbreviation. See airlines to get name.',
'flight', 'Flight number.',
'tailnum', 'Plane tail number. See planes for additional metadata.',
'origin', 'Origin and destination. See airports for additional metadata.',
'dest', 'Origin and destination. See airports for additional metadata.',
'air_time', 'Amount of time spent in the air, in minutes.',
'distance', 'Distance between airports, in miles.',
'hour', 'Time of scheduled departure broken into hour and minutes. - hour',
'minute', 'Time of scheduled departure broken into hour and minutes. minute',
'time_hour', 'Scheduled date and hour of the flight as a POSIXct date. Along with origin, can be used to
)

kable(nycflights13_names)
```

columns	description
year	Date of departure. - year
month	Date of departure. - month
day	Date of departure.
dep_time	Actual departure and arrival times (format HHMM or HMM), local tz.
arr_time	Actual departure and arrival times (format HHMM or HMM), local tz.
sched_dep_time	Scheduled departure and arrival times (format HHMM or HMM), local tz.
sched_arr_time	Scheduled departure and arrival times (format HHMM or HMM), local tz.
dep_delay	Departure and arrival delays, in minutes. Negative times represent early departures/
arr_delay	Departure and arrival delays, in minutes. Negative times represent early departures/
carrier	Two letter carrier abbreviation. See airlines to get name.
flight	Flight number.
tailnum	Plane tail number. See planes for additional metadata.
origin	Origin and destination. See airports for additional metadata.
dest	Origin and destination. See airports for additional metadata.
air_time	Amount of time spent in the air, in minutes.
distance	Distance between airports, in miles.
hour	Time of scheduled departure broken into hour and minutes. - hour
minute	Time of scheduled departure broken into hour and minutes. minute
time_hour	Scheduled date and hour of the flight as a POSIXct date. Along with origin, can be used to join fligh
reading the Readme	file

```
X <- read_html("/Users/JaggerCes/Documents/Grad School/Fall 19/CSUEB/Stats 650 - Adv_R_for_Data_Sci/sta
X
```

```
##   [1] "BACKGROUND"
##   [2] "The data contained in the compressed file has been extracted from the \n\tReporting Carrier On
##   [3] "RECORD LAYOUT"
##   [4] "Below are fields in the order that they appear on the records:"
##   [5] "Year"
##   [6] "Year"
##   [7] "Quarter"
##   [8] "Quarter (1-4)"
##   [9] "Month"
##  [10] "Month"
##  [11] "DayofMonth"
##  [12] "Day of Month"
##  [13] "DayOfWeek"
##  [14] "Day of Week"
##  [15] "FlightDate"
##  [16] "Flight Date (yyyymmdd)"
##  [17] "Reporting_Airline"
##  [18] "Unique Carrier Code. When the same code has been used by multiple carriers, a numeric suffix i
##  [19] "DOT_ID_Reporting_Airline"
##  [20] "An identification number assigned by US DOT to identify a unique airline (carrier). A unique a
##  [21] "IATA_CODE_Reporting_Airline"
##  [22] "Code assigned by IATA and commonly used to identify a carrier. As the same code may have been
##  [23] "Tail_Number"
##  [24] "Tail Number"
##  [25] "Flight_Number_Reporting_Airline"
##  [26] "Flight Number"
##  [27] "OriginAirportID"
##  [28] "Origin Airport, Airport ID. An identification number assigned by US DOT to identify a unique a
##  [29] "OriginAirportSeqID"
##  [30] "Origin Airport, Airport Sequence ID. An identification number assigned by US DOT to identify a
##  [31] "OriginCityMarketID"
##  [32] "Origin Airport, City Market ID. City Market ID is an identification number assigned by US DOT
##  [33] "Origin"
##  [34] "Origin Airport"
##  [35] "OriginCityName"
##  [36] "Origin Airport, City Name"
##  [37] "OriginState"
##  [38] "Origin Airport, State Code"
##  [39] "OriginStateFips"
##  [40] "Origin Airport, State Fips"
##  [41] "OriginStateName"
##  [42] "Origin Airport, State Name"
##  [43] "OriginWac"
##  [44] "Origin Airport, World Area Code"
##  [45] "DestAirportID"
##  [46] "Destination Airport, Airport ID. An identification number assigned by US DOT to identify a uni
##  [47] "DestAirportSeqID"
##  [48] "Destination Airport, Airport Sequence ID. An identification number assigned by US DOT to ident
##  [49] "DestCityMarketID"
##  [50] "Destination Airport, City Market ID. City Market ID is an identification number assigned by US
##  [51] "Dest"
##  [52] "Destination Airport"
##  [53] "DestCityName"
##  [54] "Destination Airport, City Name"
```

```
##  [55] "DestState"
##  [56] "Destination Airport, State Code"
##  [57] "DestStateFips"
##  [58] "Destination Airport, State Fips"
##  [59] "DestStateName"
##  [60] "Destination Airport, State Name"
##  [61] "DestWac"
##  [62] "Destination Airport, World Area Code"
##  [63] "CRSDepTime"
##  [64] "CRS Departure Time (local time: hhmm)"
##  [65] "DepTime"
##  [66] "Actual Departure Time (local time: hhmm)"
##  [67] "DepDelay"
##  [68] "Difference in minutes between scheduled and actual departure time. Early departures show nega
##  [69] "DepDelayMinutes"
##  [70] "Difference in minutes between scheduled and actual departure time. Early departures set to 0.
##  [71] "DepDel15"
##  [72] "Departure Delay Indicator, 15 Minutes or More (1=Yes)"
##  [73] "DepartureDelayGroups"
##  [74] "Departure Delay intervals, every (15 minutes from 180)"
##  [75] "DepTimeBlk"
##  [76] "CRS Departure Time Block, Hourly Intervals"
##  [77] "TaxiOut"
##  [78] "Taxi Out Time, in Minutes"
##  [79] "WheelsOff"
##  [80] "Wheels Off Time (local time: hhmm)"
##  [81] "WheelsOn"
##  [82] "Wheels On Time (local time: hhmm)"
##  [83] "TaxiIn"
##  [84] "Taxi In Time, in Minutes"
##  [85] "CRSArrTime"
##  [86] "CRS Arrival Time (local time: hhmm)"
##  [87] "ArrTime"
##  [88] "Actual Arrival Time (local time: hhmm)"
##  [89] "ArrDelay"
##  [90] "Difference in minutes between scheduled and actual arrival time. Early arrivals show negative
##  [91] "ArrDelayMinutes"
##  [92] "Difference in minutes between scheduled and actual arrival time. Early arrivals set to 0."
##  [93] "ArrDel15"
##  [94] "Arrival Delay Indicator, 15 Minutes or More (1=Yes)"
##  [95] "ArrivalDelayGroups"
##  [96] "Arrival Delay intervals, every (15-minutes from 180)"
##  [97] "ArrTimeBlk"
##  [98] "CRS Arrival Time Block, Hourly Intervals"
##  [99] "Cancelled"
## [100] "Cancelled Flight Indicator (1=Yes)"
## [101] "CancellationCode"
## [102] "Specifies The Reason For Cancellation"
## [103] "Diverted"
## [104] "Diverted Flight Indicator (1=Yes)"
## [105] "CRSElapsedTime"
## [106] "CRS Elapsed Time of Flight, in Minutes"
## [107] "ActualElapsedTime"
## [108] "Elapsed Time of Flight, in Minutes"
```

```
## [109] "AirTime"
## [110] "Flight Time, in Minutes"
## [111] "Flights"
## [112] "Number of Flights"
## [113] "Distance"
## [114] "Distance between airports (miles)"
## [115] "DistanceGroup"
## [116] "Distance Intervals, every 250 Miles, for Flight Segment"
## [117] "CarrierDelay"
## [118] "Carrier Delay, in Minutes"
## [119] "WeatherDelay"
## [120] "Weather Delay, in Minutes"
## [121] "NASDelay"
## [122] "National Air System Delay, in Minutes"
## [123] "SecurityDelay"
## [124] "Security Delay, in Minutes"
## [125] "LateAircraftDelay"
## [126] "Late Aircraft Delay, in Minutes"
## [127] "FirstDepTime"
## [128] "First Gate Departure Time at Origin Airport"
## [129] "TotalAddGTime"
## [130] "Total Ground Time Away from Gate for Gate Return or Cancelled Flight"
## [131] "LongestAddGTime"
## [132] "Longest Time Away from Gate for Gate Return or Cancelled Flight"
## [133] "DivAirportLandings"
## [134] "Number of Diverted Airport Landings"
## [135] "DivReachedDest"
## [136] "Diverted Flight Reaching Scheduled Destination Indicator (1=Yes)"
## [137] "DivActualElapsedTime"
## [138] "Elapsed Time of Diverted Flight Reaching Scheduled Destination, in Minutes.  The ActualElapsed
## [139] "DivArrDelay"
## [140] "Difference in minutes between scheduled and actual arrival time for a diverted flight reaching
## [141] "DivDistance"
## [142] "Distance between scheduled destination and final diverted airport (miles).  Value will be 0 fo
## [143] "Div1Airport"
## [144] "Diverted Airport Code1"
## [145] "Div1AirportID"
## [146] "Airport ID of Diverted Airport 1. Airport ID is a Unique Key for an Airport"
## [147] "Div1AirportSeqID"
## [148] "Airport Sequence ID of Diverted Airport 1. Unique Key for Time Specific Information for an Air
## [149] "Div1WheelsOn"
## [150] "Wheels On Time (local time: hhmm) at Diverted Airport Code1"
## [151] "Div1TotalGTime"
## [152] "Total Ground Time Away from Gate at Diverted Airport Code1"
## [153] "Div1LongestGTime"
## [154] "Longest Ground Time Away from Gate at Diverted Airport Code1"
## [155] "Div1WheelsOff"
## [156] "Wheels Off Time (local time: hhmm) at Diverted Airport Code1"
## [157] "Div1TailNum"
## [158] "Aircraft Tail Number for Diverted Airport Code1"
## [159] "Div2Airport"
## [160] "Diverted Airport Code2"
## [161] "Div2AirportID"
## [162] "Airport ID of Diverted Airport 2. Airport ID is a Unique Key for an Airport"
```

```
## [163] "Div2AirportSeqID"
## [164] "Airport Sequence ID of Diverted Airport 2. Unique Key for Time Specific Information for an Ai
## [165] "Div2WheelsOn"
## [166] "Wheels On Time (local time: hhmm) at Diverted Airport Code2"
## [167] "Div2TotalGTime"
## [168] "Total Ground Time Away from Gate at Diverted Airport Code2"
## [169] "Div2LongestGTime"
## [170] "Longest Ground Time Away from Gate at Diverted Airport Code2"
## [171] "Div2WheelsOff"
## [172] "Wheels Off Time (local time: hhmm) at Diverted Airport Code2"
## [173] "Div2TailNum"
## [174] "Aircraft Tail Number for Diverted Airport Code2"
## [175] "Div3Airport"
## [176] "Diverted Airport Code3"
## [177] "Div3AirportID"
## [178] "Airport ID of Diverted Airport 3. Airport ID is a Unique Key for an Airport"
## [179] "Div3AirportSeqID"
## [180] "Airport Sequence ID of Diverted Airport 3. Unique Key for Time Specific Information for an Ai
## [181] "Div3WheelsOn"
## [182] "Wheels On Time (local time: hhmm) at Diverted Airport Code3"
## [183] "Div3TotalGTime"
## [184] "Total Ground Time Away from Gate at Diverted Airport Code3"
## [185] "Div3LongestGTime"
## [186] "Longest Ground Time Away from Gate at Diverted Airport Code3"
## [187] "Div3WheelsOff"
## [188] "Wheels Off Time (local time: hhmm) at Diverted Airport Code3"
## [189] "Div3TailNum"
## [190] "Aircraft Tail Number for Diverted Airport Code3"
## [191] "Div4Airport"
## [192] "Diverted Airport Code4"
## [193] "Div4AirportID"
## [194] "Airport ID of Diverted Airport 4. Airport ID is a Unique Key for an Airport"
## [195] "Div4AirportSeqID"
## [196] "Airport Sequence ID of Diverted Airport 4. Unique Key for Time Specific Information for an Ai
## [197] "Div4WheelsOn"
## [198] "Wheels On Time (local time: hhmm) at Diverted Airport Code4"
## [199] "Div4TotalGTime"
## [200] "Total Ground Time Away from Gate at Diverted Airport Code4"
## [201] "Div4LongestGTime"
## [202] "Longest Ground Time Away from Gate at Diverted Airport Code4"
## [203] "Div4WheelsOff"
## [204] "Wheels Off Time (local time: hhmm) at Diverted Airport Code4"
## [205] "Div4TailNum"
## [206] "Aircraft Tail Number for Diverted Airport Code4"
## [207] "Div5Airport"
## [208] "Diverted Airport Code5"
## [209] "Div5AirportID"
## [210] "Airport ID of Diverted Airport 5. Airport ID is a Unique Key for an Airport"
## [211] "Div5AirportSeqID"
## [212] "Airport Sequence ID of Diverted Airport 5. Unique Key for Time Specific Information for an Ai
## [213] "Div5WheelsOn"
## [214] "Wheels On Time (local time: hhmm) at Diverted Airport Code5"
## [215] "Div5TotalGTime"
## [216] "Total Ground Time Away from Gate at Diverted Airport Code5"
```

```
## [217] "Div5LongestGTime"
## [218] "Longest Ground Time Away from Gate at Diverted Airport Code5"
## [219] "Div5WheelsOff"
## [220] "Wheels Off Time (local time: hhmm) at Diverted Airport Code5"
## [221] "Div5TailNum"
## [222] "Aircraft Tail Number for Diverted Airport Code5"
```

```
#removing the top 4 lines
X <- X[-c(1,2,3,4)]
head(X)
```

```
## [1] "Year"          "Year"          "Quarter"       "Quarter (1-4)"
## [5] "Month"         "Month"
```

creating a matrix for the variable names in the firstcolumn and description in the second

```
Y <- matrix(X[1:218], ncol = 2, byrow = TRUE)
head(Y)
```

```
##      [,1]         [,2]
## [1,] "Year"       "Year"
## [2,] "Quarter"    "Quarter (1-4)"
## [3,] "Month"      "Month"
## [4,] "DayofMonth" "Day of Month"
## [5,] "DayOfWeek"  "Day of Week"
## [6,] "FlightDate" "Flight Date (yyyymmdd)"
```

Putting the matrix in tibble

```
on_time_names <- tibble(columns = Y[,1], description = Y[,2]) %>%
  mutate(columns = tolower(columns))

head(on_time_names)
```

```
## # A tibble: 6 x 2
##   columns    description
##   <chr>      <chr>
## 1 year       Year
## 2 quarter    Quarter (1-4)
## 3 month      Month
## 4 dayofmonth Day of Month
## 5 dayofweek  Day of Week
## 6 flightdate Flight Date (yyyymmdd)
```

```
nycflights13_names <- nycflights13_names %>%
  mutate(columns = as_factor(columns))

#listing variables in nycflights13
nycflights13_names$columns %>% levels()
```

```
##  [1] "year"           "month"          "day"            "dep_time"
##  [5] "arr_time"       "sched_dep_time" "sched_arr_time" "dep_delay"
##  [9] "arr_delay"      "carrier"        "flight"         "tailnum"
## [13] "origin"         "dest"           "air_time"       "distance"
## [17] "hour"           "minute"         "time_hour"
```

```
#number of variables in nycflights13
nycflights13_names$columns %>% nlevels()
```

```
## [1] 19
```

Renaming variables in nycflights13

```
nycflights13_names$columns <- nycflights13_names$columns %>%
  fct_recode("dayofmonth" = "day",
             "deptime" = "dep_time",
             "arrtime" = "arr_time",
             "crsdeptime" = "sched_dep_time",
             "crsarrtime" = "sched_arr_time",
             "depdelay" = "dep_delay",
             "arrdelay" = "arr_delay",
             "reporting_airline" = "carrier",
             "flight_number_reporting_airline" = "flight",
             "tail_number"= "tailnum",
             "airtime" = "air_time",
             "crsdeptime" = "time_hour"
)

head(nycflights13_names)
```

```
## # A tibble: 6 x 2
##   columns    description
##   <fct>      <chr>
## 1 year       Date of departure. - year
## 2 month      Date of departure. - month
## 3 dayofmonth Date of departure.
## 4 deptime    Actual departure and arrival times (format HHMM or HMM), loca~
## 5 arrtime    Actual departure and arrival times (format HHMM or HMM), loca~
## 6 crsdeptime Scheduled departure and arrival times (format HHMM or HMM), l~
```

3. Matching variables.

```
match <- inner_join(nycflights13_names, on_time_names, "columns")
```

```
## Warning: Column `columns` joining factor and character vector, coercing
## into character vector
```

```
head(match)
```

```
## # A tibble: 6 x 3
##   columns   description.x                          description.y
##   <chr>     <chr>                                  <chr>
## 1 year      Date of departure. - year              Year
## 2 month     Date of departure. - month             Month
## 3 dayofmon~ Date of departure.                     Day of Month
## 4 deptime   Actual departure and arrival times (f~ Actual Departure Time (~
## 5 arrtime   Actual departure and arrival times (f~ Actual Arrival Time (lo~
## 6 crsdepti~ Scheduled departure and arrival times~ CRS Departure Time (loc~
```

Variables not in the On-Time data.

```
anti_join(nycflights13_names, on_time_names, "columns")
```

```
## Warning: Column `columns` joining factor and character vector, coercing
## into character vector
```

```
## # A tibble: 2 x 2
##   columns description
```

```
##   <fct>   <chr>
## 1 hour    Time of scheduled departure broken into hour and minutes. - hour
## 2 minute  Time of scheduled departure broken into hour and minutes. minute
```

Question 4: Variables not in the nycflights13 data

```
new_vars <- anti_join(on_time_names, nycflights13_names, "columns")
```

```
## Warning: Column `columns` joining character vector and factor, coercing
## into character vector
```

```
new_vars_tibbles <- tibble(columns = new_vars$columns, description = new_vars$description)
head(new_vars_tibbles, 10)
```

```
## # A tibble: 10 x 2
##    columns               description
##    <chr>                 <chr>
##  1 quarter               Quarter (1-4)
##  2 dayofweek             Day of Week
##  3 flightdate            Flight Date (yyyymmdd)
##  4 dot_id_reporting_air~ An identification number assigned by US DOT to id~
##  5 iata_code_reporting_~ Code assigned by IATA and commonly used to identi~
##  6 originairportid       Origin Airport, Airport ID. An identification num~
##  7 originairportseqid    Origin Airport, Airport Sequence ID. An identific~
##  8 origincitymarketid    Origin Airport, City Market ID. City Market ID is~
##  9 origincityname        Origin Airport, City Name
## 10 originstate           Origin Airport, State Code
```

Problem 5: Excercise 4.2 The data provides two ways of checking for cancelled flights. Cancelled and missing data on arrival (ArrDelay). Yet Cancelled reports that no flights that departed from either SFO, OAK, or SJC were cancelled in Feb, Apr, Jun, Sep, Oct, and Nov 2018. This seems very unusual, which is why I'd use the missing data in ArrDelay. Thus August had the highest proportion of cancelled fights, and February had the lowest. This suggests that both the volume and weather have a big effect on cancelled flights in the Bay-Area.

```
sfoflights18 <- my_flights %>%
  filter(Origin == "SFO" | Origin == "OAK" | Origin == "SJC")
head(sfoflights18)
```

```
## # A tibble: 6 x 110
##    Year Quarter Month DayofMonth DayOfWeek FlightDate Reporting_Airli~
##   <dbl>   <dbl> <dbl>      <dbl>     <dbl> <date>     <chr>
## 1  2018       1     1         27         6 2018-01-27 UA
## 2  2018       1     1         27         6 2018-01-27 UA
## 3  2018       1     1         27         6 2018-01-27 UA
## 4  2018       1     1         27         6 2018-01-27 UA
## 5  2018       1     1         27         6 2018-01-27 UA
## 6  2018       1     1         27         6 2018-01-27 UA
## # ... with 103 more variables: DOT_ID_Reporting_Airline <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, Tail_Number <chr>,
## #   Flight_Number_Reporting_Airline <dbl>, OriginAirportID <dbl>,
## #   OriginAirportSeqID <dbl>, OriginCityMarketID <dbl>, Origin <chr>,
## #   OriginCityName <chr>, OriginState <chr>, OriginStateFips <chr>,
## #   OriginStateName <chr>, OriginWac <dbl>, DestAirportID <dbl>,
## #   DestAirportSeqID <dbl>, DestCityMarketID <dbl>, Dest <chr>,
## #   DestCityName <chr>, DestState <chr>, DestStateFips <chr>,
## #   DestStateName <chr>, DestWac <dbl>, CRSDepTime <chr>, DepTime <chr>,
```

```
## #   DepDelay <dbl>, DepDelayMinutes <dbl>, DepDel15 <dbl>,
## #   DepartureDelayGroups <dbl>, DepTimeBlk <chr>, TaxiOut <dbl>,
## #   WheelsOff <chr>, WheelsOn <chr>, TaxiIn <dbl>, CRSArrTime <chr>,
## #   ArrTime <chr>, ArrDelay <dbl>, ArrDelayMinutes <dbl>, ArrDel15 <dbl>,
## #   ArrivalDelayGroups <dbl>, ArrTimeBlk <chr>, Cancelled <dbl>,
## #   CancellationCode <chr>, Diverted <dbl>, CRSElapsedTime <dbl>,
## #   ActualElapsedTime <dbl>, AirTime <dbl>, Flights <dbl>, Distance <dbl>,
## #   DistanceGroup <dbl>, CarrierDelay <dbl>, WeatherDelay <dbl>,
## #   NASDelay <dbl>, SecurityDelay <dbl>, LateAircraftDelay <dbl>,
## #   FirstDepTime <dbl>, TotalAddGTime <dbl>, LongestAddGTime <dbl>,
## #   DivAirportLandings <dbl>, DivReachedDest <dbl>,
## #   DivActualElapsedTime <dbl>, DivArrDelay <dbl>, DivDistance <dbl>,
## #   Div1Airport <chr>, Div1AirportID <dbl>, Div1AirportSeqID <dbl>,
## #   Div1WheelsOn <dbl>, Div1TotalGTime <dbl>, Div1LongestGTime <dbl>,
## #   Div1WheelsOff <dbl>, Div1TailNum <chr>, Div2Airport <lgl>,
## #   Div2AirportID <lgl>, Div2AirportSeqID <lgl>, Div2WheelsOn <lgl>,
## #   Div2TotalGTime <lgl>, Div2LongestGTime <lgl>, Div2WheelsOff <lgl>,
## #   Div2TailNum <lgl>, Div3Airport <lgl>, Div3AirportID <lgl>,
## #   Div3AirportSeqID <lgl>, Div3WheelsOn <lgl>, Div3TotalGTime <lgl>,
## #   Div3LongestGTime <lgl>, Div3WheelsOff <lgl>, Div3TailNum <lgl>,
## #   Div4Airport <lgl>, Div4AirportID <lgl>, Div4AirportSeqID <lgl>,
## #   Div4WheelsOn <lgl>, Div4TotalGTime <lgl>, Div4LongestGTime <lgl>,
## #   Div4WheelsOff <lgl>, Div4TailNum <lgl>, Div5Airport <lgl>,
## #   Div5AirportID <lgl>, Div5AirportSeqID <lgl>, Div5WheelsOn <lgl>,
## #   Div5TotalGTime <lgl>, Div5LongestGTime <lgl>, ...
```

```
#flights that were reported cancelled
sfoflights18 %>% group_by(Cancelled, Month) %>%
  filter(Cancelled == 1) %>%
  count() %>%
  arrange(desc(n))
```

```
## # A tibble: 12 x 3
## # Groups:   Cancelled, Month [12]
##    Cancelled Month     n
##        <dbl> <dbl> <int>
##  1         1     3   519
##  2         1     8   477
##  3         1     1   332
##  4         1     5   302
##  5         1    12   289
##  6         1     7   280
##  7         1     4   258
##  8         1    11   254
##  9         1     6   246
## 10         1     9   230
## 11         1    10   173
## 12         1     2   144
```

```
#flights that that were missing arrival data and could thus be considered cancelled
sfoflights18 %>% group_by(ArrDelay, Month) %>%
  filter(is.na(ArrDelay)) %>%
  count() %>%
  arrange(desc(n))
```

```
## # A tibble: 12 x 3
```

```
## # Groups:   ArrDelay, Month [12]
##    ArrDelay Month     n
##       <dbl> <dbl> <int>
##  1       NA     8   565
##  2       NA     3   560
##  3       NA     1   391
##  4       NA     5   357
##  5       NA     7   350
##  6       NA    12   341
##  7       NA     6   306
##  8       NA     4   304
##  9       NA    11   302
## 10       NA     9   280
## 11       NA    10   235
## 12       NA     2   184
```

Problem 5: Excercise 4.3 Tail number of the plan with the most flights from the Bay-Area N286VA.

```
sfoflights18 %>% select(Tail_Number) %>%
  group_by(Tail_Number) %>%
  count() %>%
  arrange(desc(n))
```

```
## # A tibble: 3,983 x 2
## # Groups:   Tail_Number [3,983]
##    Tail_Number     n
##    <chr>       <int>
##  1 N286VA        406
##  2 N284VA        401
##  3 N849VA        394
##  4 N630VA        390
##  5 N632VA        388
##  6 N528VA        387
##  7 N636VA        383
##  8 N635VA        381
##  9 N283VA        378
## 10 N526VA        375
## # ... with 3,973 more rows
```

Plot the number of trips per week over the year.

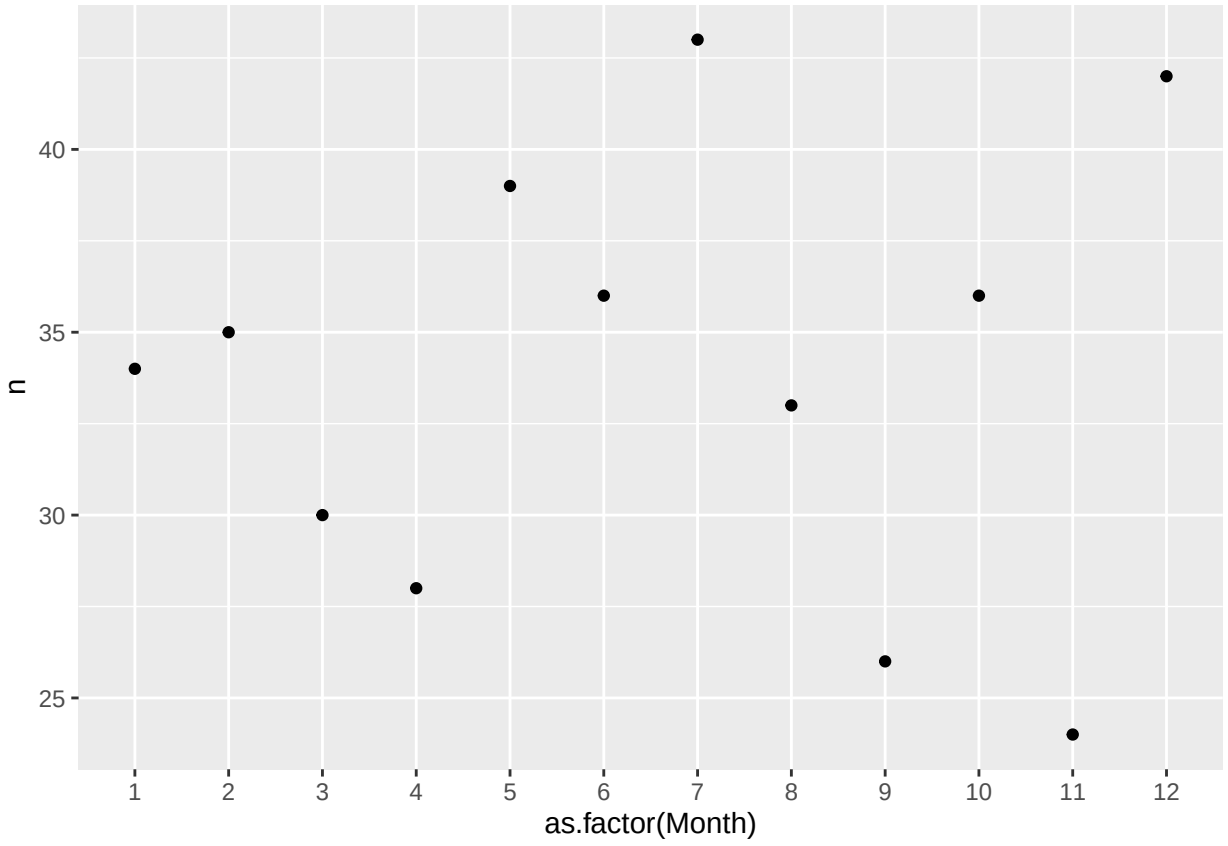
```
sfoflights18 %>% select(Year, Month, DayofMonth, Tail_Number) %>%
  filter(Tail_Number == "N286VA") %>%
  group_by(Year, Month) %>%
  summarize(count=n())
```

```
## # A tibble: 12 x 3
## # Groups:   Year [1]
##     Year Month count
##    <dbl> <dbl> <int>
##  1  2018     1    34
##  2  2018     2    35
##  3  2018     3    30
##  4  2018     4    28
##  5  2018     5    39
##  6  2018     6    36
```

```
##  7  2018     7    43
##  8  2018     8    33
##  9  2018     9    26
## 10  2018    10    36
## 11  2018    11    24
## 12  2018    12    42
```

```
sfoflights18 %>% filter(Tail_Number == "N286VA") %>%
  group_by(Month) %>%
  tally() %>%
  ggplot(aes(x = as.factor(Month), y = n)) +
  geom_point() +
  geom_line()
```

```
## geom_path: Each group consists of only one observation. Do you need to
## adjust the group aesthetic?
```



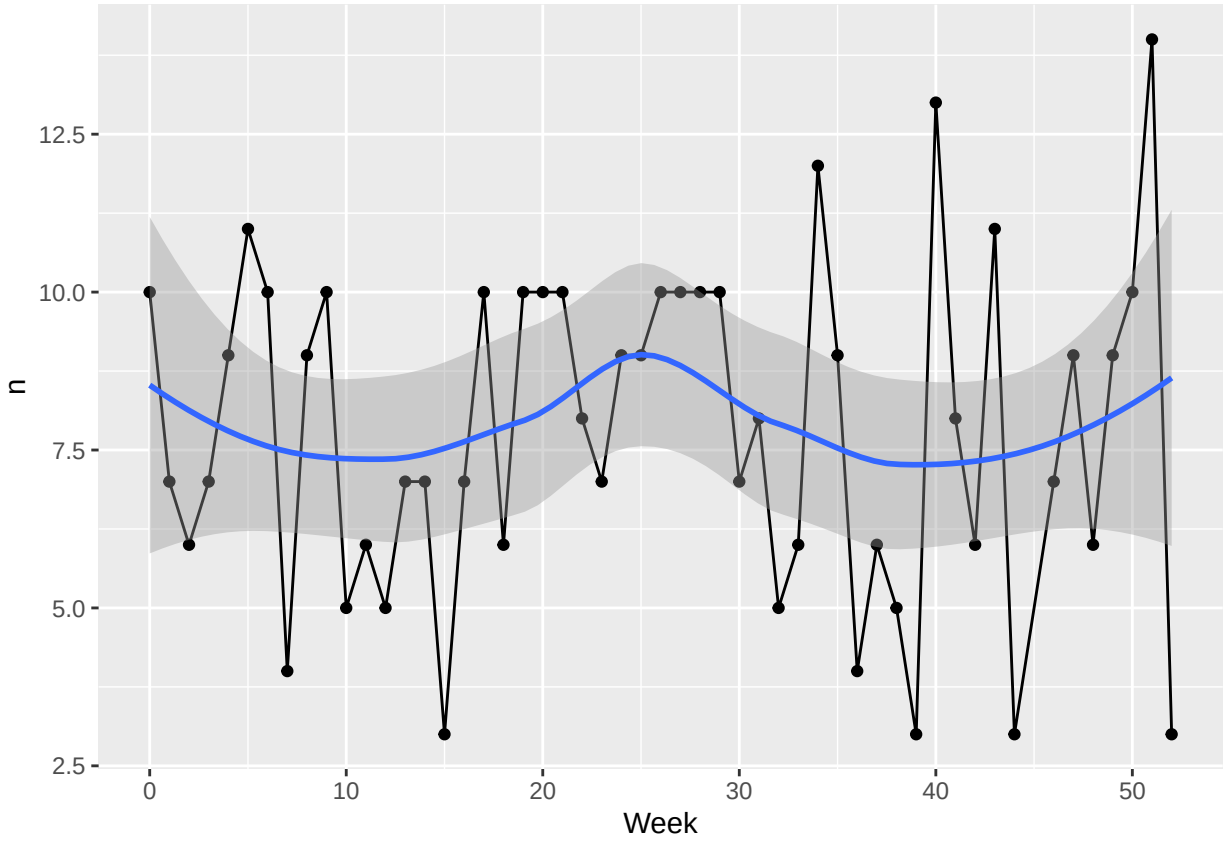

```
flights2 <- sfoflights18 %>% mutate( ymd = as.Date(paste(Year, Month, DayofMonth, sep='-'))) %>%
  mutate( Week = as.integer(format(as.Date(ymd), "%U"))  ) %>%
  filter(Tail_Number == "N286VA") %>%
  group_by(Week) %>%
  tally()

flights2
```

```
## # A tibble: 52 x 2
##     Week     n
##    <int> <int>
```

```
##  1     0    10
##  2     1     7
##  3     2     6
##  4     3     7
##  5     4     9
##  6     5    11
##  7     6    10
##  8     7     4
##  9     8     9
## 10     9    10
## # ... with 42 more rows
```

```
flights2 %>% ggplot(aes(x=Week, y=n)) +
  geom_point() +
  geom_line() +
  geom_smooth()
```

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



Problem 5: Excercise 4.4 The oldest plane flying from the Bay-Area in 2018 was the N505UA and it was manufactured in 1989.

```
planes2 <- planes %>% rename( year_m = year )

planes2
```

```
## # A tibble: 3,322 x 9
##    tailnum year_m type      manufacturer  model  engines seats speed engine
##    <chr>    <int> <chr>     <chr>         <chr>    <int> <int> <int> <chr>
```

```
##  1 N10156    2004 Fixed wi~ EMBRAER       EMB-1~       2    55    NA Turbo~
##  2 N102UW    1998 Fixed wi~ AIRBUS INDUS~ A320-~       2   182    NA Turbo~
##  3 N103US    1999 Fixed wi~ AIRBUS INDUS~ A320-~       2   182    NA Turbo~
##  4 N104UW    1999 Fixed wi~ AIRBUS INDUS~ A320-~       2   182    NA Turbo~
##  5 N10575    2002 Fixed wi~ EMBRAER       EMB-1~       2    55    NA Turbo~
##  6 N105UW    1999 Fixed wi~ AIRBUS INDUS~ A320-~       2   182    NA Turbo~
##  7 N107US    1999 Fixed wi~ AIRBUS INDUS~ A320-~       2   182    NA Turbo~
##  8 N108UW    1999 Fixed wi~ AIRBUS INDUS~ A320-~       2   182    NA Turbo~
##  9 N109UW    1999 Fixed wi~ AIRBUS INDUS~ A320-~       2   182    NA Turbo~
## 10 N110UW    1999 Fixed wi~ AIRBUS INDUS~ A320-~       2   182    NA Turbo~
## # ... with 3,312 more rows
```

```
planes2 %>% inner_join(sfoflights18, by = c("tailnum" = "Tail_Number")) %>%
  group_by(tailnum, year_m, Year) %>%
  count() %>%
  arrange(year_m)
```

```
## # A tibble: 1,956 x 4
## # Groups:   tailnum, year_m, Year [1,956]
##    tailnum year_m  Year     n
##    <chr>    <int> <dbl> <int>
##  1 N502UA    1989  2018   217
##  2 N505UA    1989  2018   228
##  3 N152DL    1990  2018     6
##  4 N171DN    1990  2018     5
##  5 N172DN    1990  2018     5
##  6 N174DN    1990  2018     5
##  7 N175DN    1990  2018     7
##  8 N176DN    1990  2018     3
##  9 N309US    1990  2018    13
## 10 N312US    1990  2018     2
## # ... with 1,946 more rows
```

There were 260 planes that flew from the Bay-Area and are included in the planes table?

```
C <- sfoflights18 %>% inner_join(planes2, by = c("Tail_Number" = "tailnum")) %>%
  group_by(Tail_Number) %>%
  count()

num_planes = n_distinct(C$n)
num_planes
```

```
## [1] 260
```

Taking a sample of 100,000 rows

```
sample_100k <- sfoflights18 %>%
  sample(100000:nrow(sfoflights18), 110)
sample_100k
```

```
## # A tibble: 283,150 x 100,000
##    Div1TailNum DepDelay Div1AirportSeqID CRSArrTime DepartureDelayG~
##    <chr>          <dbl>            <dbl> <chr>                 <dbl>
##  1 <NA>              -9               NA 1221                     -1
##  2 <NA>             -18               NA 2007                     -2
##  3 <NA>             -14               NA 1913                     -1
##  4 <NA>              -6               NA 2359                     -1
```

```
##  5 <NA>              -5              NA 1659                     -1
##  6 <NA>              -8              NA 1615                     -1
##  7 <NA>              -5              NA 2132                     -1
##  8 <NA>              -1              NA 1717                     -1
##  9 <NA>               8              NA 2251                      0
## 10 <NA>              -8              NA 0009                     -1
## # ... with 283,140 more rows, and 99,995 more variables:
## #   DistanceGroup <dbl>, Div2WheelsOn <lgl>, Div5AirportID <lgl>,
## #   Year <dbl>, DepTimeBlk <chr>, AirTime <dbl>, Div5AirportSeqID <lgl>,
## #   Div2AirportSeqID <lgl>, X110 <lgl>, DivAirportLandings <dbl>,
## #   OriginAirportSeqID <dbl>, DestAirportID <dbl>,
## #   OriginCityMarketID <dbl>, DivReachedDest <dbl>,
## #   DestCityMarketID <dbl>, OriginStateFips <chr>, TaxiOut <dbl>,
## #   Div3WheelsOff <lgl>, FlightDate <date>, FirstDepTime <dbl>,
## #   DestWac <dbl>, CRSDepTime <chr>, OriginWac <dbl>, DivDistance <dbl>,
## #   Flights <dbl>, FirstDepTime <dbl>, Div5TailNum <lgl>,
## #   OriginStateFips <chr>, DestWac <dbl>, WheelsOn <chr>, DepDelay <dbl>,
## #   Div2LongestGTime <lgl>, DestCityName <chr>, DepDelay <dbl>,
## #   X110 <lgl>, Origin <chr>, OriginWac <dbl>, OriginStateName <chr>,
## #   DestAirportSeqID <dbl>, Div2AirportSeqID <lgl>,
## #   ActualElapsedTime <dbl>, Month <dbl>, Div1WheelsOn <dbl>,
## #   OriginWac <dbl>, Month <dbl>, Div2WheelsOn <lgl>, Div1AirportID <dbl>,
## #   Div1LongestGTime <dbl>, DestStateFips <chr>, FirstDepTime <dbl>,
## #   Div5WheelsOn <lgl>, Div4TailNum <lgl>, Div1TailNum <chr>,
## #   LongestAddGTime <dbl>, Year <dbl>, OriginStateFips <chr>,
## #   Div4Airport <lgl>, LongestAddGTime <dbl>, Reporting_Airline <chr>,
## #   Div1LongestGTime <dbl>, DestAirportSeqID <dbl>, DistanceGroup <dbl>,
## #   Div1AirportSeqID <dbl>, Div4TotalGTime <lgl>, Div5Airport <lgl>,
## #   Div5TotalGTime <lgl>, Div2WheelsOn <lgl>, OriginAirportSeqID <dbl>,
## #   OriginCityMarketID <dbl>, SecurityDelay <dbl>, Div4Airport <lgl>,
## #   Div5TotalGTime <lgl>, Div3AirportID <lgl>, LongestAddGTime <dbl>,
## #   NASDelay <dbl>, Year <dbl>, DestStateFips <chr>, DepDel15 <dbl>,
## #   Reporting_Airline <chr>, OriginAirportID <dbl>, Div5WheelsOff <lgl>,
## #   Div2LongestGTime <lgl>, Div5Airport <lgl>, DivAirportLandings <dbl>,
## #   IATA_CODE_Reporting_Airline <chr>, LateAircraftDelay <dbl>,
## #   DestStateFips <chr>, Div3AirportID <lgl>, Div1WheelsOff <dbl>,
## #   Flights <dbl>, CRSDepTime <chr>, FlightDate <date>,
## #   OriginAirportSeqID <dbl>, WheelsOn <chr>, DayOfWeek <dbl>,
## #   DestStateName <chr>, DestState <chr>, DestCityMarketID <dbl>,
## #   LongestAddGTime <dbl>, Div3WheelsOff <lgl>, ...
```