

The Big Ones

The Big Ones

The big ones, where you're going to make the big money with the best contracts, are really just Delta, United, American, FedEx, and UPS. OK, let's look at what proportion of flights in NYC for the year 2013 were one of these "big ones". "Carrier" field is the name of the airline:

- "DL" for Delta
- "UA" for United
- "AA" for American
- "FX" for Federal Express Corporation
- "5X" for United Parcel Service

First we need to import Libraries. This imports the data set of flights and tidyverse is packages for manipulating the data.

```
## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.0      v purrr  0.3.4
## v tibble  3.0.1      v dplyr  0.8.5
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

The data "flights" has many columns or variables

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

We are only interested in the carrier so let's make dataframe with just "flight" and "carrier". And we'll look at first few lines and the total number of flights in this data.

```
## # A tibble: 336,776 x 2
##   flight carrier
##   <int> <chr>
## 1  1545 UA
## 2  1714 UA
## 3  1141 AA
## 4   725 B6
## 5   461 DL
## 6  1696 UA
## 7   507 B6
## 8  5708 EV
## 9    79 B6
## 10  301 AA
## # ... with 336,766 more rows
```

```
## [1] 336776
```

Now let's find the number of flights that just the "big ones" and calculate the proportion

```
n.big.ones <- carriers %>% filter(carrier == "DL" | carrier == "UA" |  
                                carrier == "AA" | carrier == "FX" |  
                                carrier == "5X") %>% nrow()  
n.big.ones
```

```
## [1] 139504
```

```
pro.big.ones <- n.big.ones / nrow(carriers)  
pro.big.ones
```

```
## [1] 0.4142338
```

So, it looks like in 2013 in NYC there were total of 336776 flights with 139504 were the "big ones" and that gives proportion of 0.4142338

This here just list the number of flights grouped by carrier. The B6 is Bluejet. EV is ExpressJet Airlines (?). This R code might not be easy to read because it uses this thing call pipe: %>%. But basically we start with the data carriers; feed that to group_by which groups by carrier; feed that to summarize which finds the count for each carrier; finally arrange from largers. And there seems to be only 16 carriers.

```
carriers %>% group_by(carrier) %>% summarize(count=n()) %>% arrange(desc(count))
```

```
## # A tibble: 16 x 2  
##   carrier count  
##   <chr>   <int>  
## 1 UA      58665  
## 2 B6      54635  
## 3 EV      54173  
## 4 DL      48110  
## 5 AA      32729  
## 6 MQ      26397  
## 7 US      20536  
## 8 9E      18460  
## 9 WN      12275  
## 10 VX      5162  
## 11 FL      3260  
## 12 AS       714  
## 13 F9       685  
## 14 YV       601  
## 15 HA       342  
## 16 OO        32
```

So, UA, B6 and EV and even DL have about same so let's see just UA proportiong

```
n.UA <- carriers %>% filter(carrier == "UA") %>% nrow()  
n.UA
```

```
## [1] 58665
```

```
pro.UA <- n.UA / nrow(carriers)  
pro.UA
```

```
## [1] 0.1741959
```

Now take top 5 and see proportion:

```
n.top.five <- carriers %>% filter(carrier == "UA" | carrier == "B6" |  
                                carrier == "EV" | carrier == "DL" |  
                                carrier == "AA") %>% nrow()
```

```
n.top.five
```

```
## [1] 248312
```

```
pro.top.five <- n.top.five / nrow(carriers)  
pro.top.five
```

```
## [1] 0.7373209
```