

Jet Blue reach to the top

Nishan Madawanarachchi

Data

The data provided are the domestic market data reported for US carriers by Bureau of Transportation Statistics; available *here*. We got 28 csv files representing market data from 1990 to 2017.

```
# loading the required librarys
```

```
library(magrittr)
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(knitr)
```

```
#Getting the CSV files except 2017 data as it is not completed
```

```
flights <- data.frame(NULL)
```

```
for(i in 1990:2016) {
```

```
  file <- paste0("657240010_T_T100D_MARKET_US_CARRIER_ONLY_",i,"_All.csv")
```

```
  flights <- rbind(flights, read.csv(paste0("q3_data/",file)) )
```

```
}
```

```
str(flights)
```

```
## 'data.frame': 5610122 obs. of 37 variables:
```

```
## $ PASSENGERS : num 0 0 0 0 0 0 0 0 0 0 ...
```

```
## $ FREIGHT : num 0 0 395 390 3058 ...
```

```
## $ MAIL : num 0 0 3230 10 698 0 3 0 0 0 ...
```

```
## $ DISTANCE : num 94 94 2330 2257 2369 ...
```

```
## $ UNIQUE_CARRIER : Factor w/ 288 levels "5J (1)","5X",...: 45 48 46 46 45 45 46 45 45 19 ...
```

```
## $ AIRLINE_ID : int 19977 19393 20355 20355 19977 19977 20355 19977 19977 19991 ...
```

```
## $ UNIQUE_CARRIER_NAME : Factor w/ 289 levels "Aerial Transit Company",...: 49 42 51 51 49 49 51 49 4 ...
```

```
## $ UNIQUE_CARRIER_ENTITY: Factor w/ 401 levels "01220","01260",...: 29 20 5 5 29 29 5 29 29 13 ...
```

```
## $ REGION : Factor w/ 7 levels "A","D","I","L",...: 2 2 2 2 2 2 2 2 2 ...
```

```
## $ CARRIER : Factor w/ 356 levels "5X","8C","9R",...: 45 48 46 46 45 45 46 45 45 18 ...
```

```
## $ CARRIER_NAME : Factor w/ 332 levels "Aerial Transit Company",...: 50 42 52 52 50 50 52 50 5 ...
```

```
## $ CARRIER_GROUP : int 3 3 3 3 3 3 3 3 3 3 ...
```

```
## $ CARRIER_GROUP_NEW : int 3 3 3 3 3 3 3 3 3 3 ...
```

```
## $ ORIGIN_AIRPORT_ID : int 14679 14679 14679 14679 14679 14679 14679 14679 14683 14683 ...
```

```
## $ ORIGIN_AIRPORT_SEQ_ID: int 1467901 1467901 1467901 1467901 1467901 1467901 1467901 1467901 1467901 146830 ...
```

```
## $ ORIGIN_CITY_MARKET_ID: int 33570 33570 33570 33570 33570 33570 33570 33570 33214 33214 ...
```

```
## $ ORIGIN : Factor w/ 1988 levels "ABE","ABI","ABQ",...: 265 265 265 265 265 265 265 265 265 ...
```

```
## $ ORIGIN_CITY_NAME : Factor w/ 1769 levels "Abilene, TX",...: 254 254 254 254 254 254 254 254 254 ...
```

```
## $ ORIGIN_STATE_ABR : Factor w/ 53 levels "AK","AL","AR",...: 5 5 5 5 5 5 5 5 44 44 ...
```

```
## $ ORIGIN_STATE_FIPS : int 6 6 6 6 6 6 6 6 48 48 ...
```

```
## $ ORIGIN_STATE_NM : Factor w/ 53 levels "Alabama","Alaska",...: 5 5 5 5 5 5 5 5 43 43 ...
```

```
## $ ORIGIN_WAC : int 91 91 91 91 91 91 91 91 74 74 ...
```

```
## $ DEST_AIRPORT_ID : int 13891 13891 13931 14027 14100 14492 14524 14869 10140 16440 ...
```

```
## $ DEST_AIRPORT_SEQ_ID : int 1389101 1389101 1393101 1402701 1410001 1449201 1452401 1486901 1014001 1644001 ...
```

```
## $ DEST_CITY_MARKET_ID : int 32575 32575 33667 34027 34100 34492 34524 34614 30140 30423 ...
```

```
## $ DEST : Factor w/ 2058 levels "ABE","ABI","ABQ",...: 220 220 222 229 234 258 261 293 ...
```

```
## $ DEST_CITY_NAME : Factor w/ 1824 levels "Abilene, TX",...: 207 207 202 303 219 233 239 251 6 ...
```

```
## $ DEST_STATE_ABR : Factor w/ 53 levels "AK","AL","AR",...: 5 5 47 9 38 27 47 46 32 45 ...
```

```
## $ DEST_STATE_FIPS : int 6 6 51 12 42 37 51 49 35 48 ...
```

```
## $ DEST_STATE_NM      : Factor w/ 53 levels "Alabama","Alaska",...: 5 5 49 9 38 33 49 47 31 44 ...
## $ DEST_WAC           : int   91 91 38 33 23 36 38 87 86 74 ...
## $ YEAR               : int   1990 1990 1990 1990 1990 1990 1990 1990 1990 1990 ...
## $ QUARTER            : int    4 4 4 4 4 4 4 4 4 4 ...
## $ MONTH              : int   10 10 10 10 10 10 10 10 10 10 ...
## $ DISTANCE_GROUP     : int    1 1 5 5 5 5 5 2 2 1 ...
## $ CLASS              : Factor w/ 4 levels "F","G","L","P": 1 1 1 1 1 1 1 1 1 1 ...
## $ X                  : logi   NA NA NA NA NA NA ...
```

Dataframe `flights` has 5610122 rows and 37 variables.

Cleaning up data for unwanted columns

```
cols <- c("YEAR", "MONTH", "PASSENGERS", "FREIGHT", "MAIL", "CARRIER_GROUP_NEW", "UNIQUE_CARRIER_NAME")
flights2 <- flights[cols]
kable(head(flights2))
```

YEAR	MONTH	PASSENGERS	FREIGHT	MAIL	CARRIER_GROUP_NEW	UNIQUE_CARRIER_NAME
1990	10	0	0	0	3	United Air Lines Inc.
1990	10	0	0	0	3	Southwest Airlines Co.
1990	10	0	395	3230	3	US Airways Inc.
1990	10	0	390	10	3	US Airways Inc.
1990	10	0	3058	698	3	United Air Lines Inc.
1990	10	0	0	0	3	United Air Lines Inc.

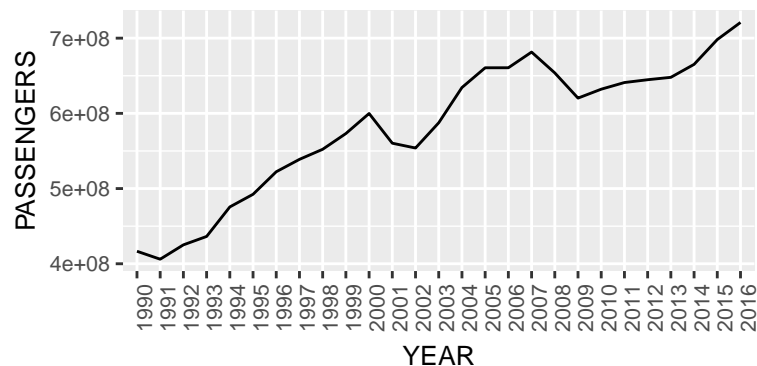
Lets first see how enplaned number of passengers grew over the years

```
# making year a factor
flights2$YEAR <- as.factor(flights2$YEAR)

# summing enplaned passengers, freight and mail per year basis
byyear_pfm <- flights2 %>% group_by(YEAR) %>%
  summarise(PASSENGERS = sum(PASSENGERS, na.rm = T), FREIGHT = sum(FREIGHT, na.rm = T),
            MAIL = sum(MAIL, na.rm = T))
kable(head(flights2))
```

YEAR	MONTH	PASSENGERS	FREIGHT	MAIL	CARRIER_GROUP_NEW	UNIQUE_CARRIER_NAME
1990	10	0	0	0	3	United Air Lines Inc.
1990	10	0	0	0	3	Southwest Airlines Co.
1990	10	0	395	3230	3	US Airways Inc.
1990	10	0	390	10	3	US Airways Inc.
1990	10	0	3058	698	3	United Air Lines Inc.
1990	10	0	0	0	3	United Air Lines Inc.

```
# plot enplaned passenger growth
ggplot(byyear_pfm, aes(x = YEAR, y = PASSENGERS, group = 1)) + geom_line() +
  theme(text = element_text(size=10), axis.text.x = element_text(angle=90))
```



Enplaned number of passengers has steadily grown over the years from 1990 to 2016. However, there are decrease in the number of passengers in the period 2000-2002 as well as in the period 2007-2009.

Lets see the mix of the carrier groups interms of number of passengers in 2016

```
# making CARRIER_GROUP_NEW a factor variable
flights2$CARRIER_GROUP_NEW <- as.factor(flights2$CARRIER_GROUP_NEW)
# looking at 2016 carrier group mix
carriergrp_2016 <- flights2 %>% filter(YEAR == 2016, PASSENGERS > 0) %>%
  group_by(CARRIER_GROUP_NEW) %>% summarise(PASSENGERS = sum(PASSENGERS,na.rm = T))

carriergrp_2016 %<>% mutate(Share = round(PASSENGERS*100/sum(PASSENGERS), 2) )
kable(carriergrp_2016)
```

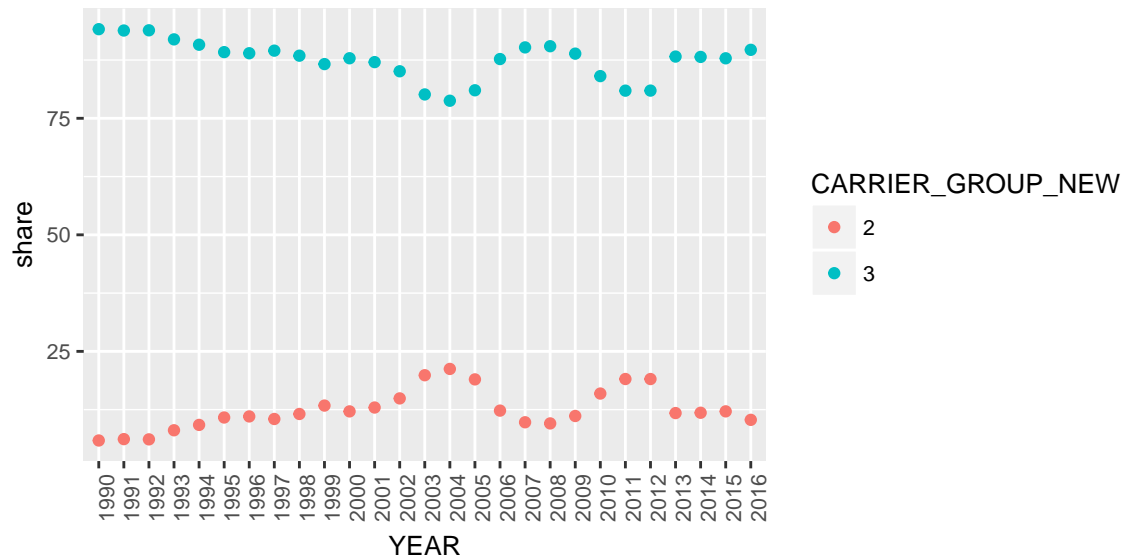
CARRIER_GROUP_NEW	PASSENGERS	Share
1	808243	0.11
2	73079382	10.14
3	635544824	88.16
4	49224	0.01
5	5068658	0.70
6	6358714	0.88
9	1643	0.00

From the table above it can be seen that Major Carriers (3) has highest share (88.16%) of enplaned passengers. Well behind major carriers are the national carriers (2) with a share of (10.14%)

Lets see how the percentage share between Major and national carriers changed over the years

```
carriergrp <- flights2 %>% filter(CARRIER_GROUP_NEW == 2 | CARRIER_GROUP_NEW == 3,
                                PASSENGERS > 0) %>%
  group_by(YEAR, CARRIER_GROUP_NEW) %>% summarise(PASSENGERS = sum(PASSENGERS,na.rm = T))

# calculating share between national carriers and major carriers
carriergrp %<>% mutate(share = round(PASSENGERS*100/sum(PASSENGERS), 2) )
ggplot(carriergrp, aes(x = YEAR, y = share, col = CARRIER_GROUP_NEW, group = 1)) +
  geom_point() + theme(text = element_text(size=10),axis.text.x = element_text(angle=90))
```



```
# maximum passenger share of national carriers
maxshare <- max(filter(carriergrp, CARRIER_GROUP_NEW==2)$share)
```

It can be seen from the graph above that national carriers were making ground in terms of passenger share in the period 1990 to 2004, where it reached the maximum passenger share of 21.23 compared to major carriers. Afterwards, the share of national carriers hits a downward trend. However, it can be assumed that during the period 2004-2008 some of the national carriers were performing well in terms of revenue and reclassified as major carriers.

Lets check for years where national carriers were reclassified as major carriers

```
# data frame of national and major carriers
natMajflights <- flights2 %>% select(YEAR, MONTH, UNIQUE_CARRIER_NAME,
                                   CARRIER_GROUP_NEW) %>%
  filter(CARRIER_GROUP_NEW == 2 | CARRIER_GROUP_NEW == 3, MONTH == 1)

# table with unique group and carrier values
natMajflights2 <- natMajflights %>% select(UNIQUE_CARRIER_NAME, CARRIER_GROUP_NEW) %>%
  distinct()

# changed list carriers
changedlist <- natMajflights2[duplicated(natMajflights2$UNIQUE_CARRIER_NAME), ]
# we are only interested in 2 to 3
changedlist <- as.character(changedlist[changedlist$CARRIER_GROUP_NEW == 3,1])

#reducing natMajflights to only group 3 carriers
natMajflights <- dplyr::filter(natMajflights, CARRIER_GROUP_NEW == 3)
indices <- match(changedlist, natMajflights$UNIQUE_CARRIER_NAME)

# Table of years when recalcification happened
kable(data.frame(Carrier = changedlist, year = natMajflights$YEAR[indices] ))
```

Carrier	year
Envoy Air	2000
ATA Airlines d/b/a ATA	2000

Carrier	year
ExpressJet Airlines Inc. (1)	2004
Comair Inc.	2005
JetBlue Airways	2005
SkyWest Airlines Inc.	2006
ExpressJet Airlines Inc.	2006
AirTran Airways Corporation	2006
Mesa Airlines Inc.	2007
Frontier Airlines Inc.	2007
Atlas Air Inc.	2008
Hawaiian Airlines Inc.	2009
Kalitta Air LLC	2012
Spirit Air Lines	2013
Virgin America	2013
Polar Air Cargo Airways	2015
Allegiant Air	2016

It is clearly visible from the above table that 9 out of the national carriers were reclassified as Major carriers during the period 2004-2008 indicating strong growth in these carriers.

Lets see the most popular Major carriers among the passengers in last 5 years

I am assuming the most popular carriers are the ones with highest enplaned passengers

```
# sum of passengers by carrier
top5 <- flights2 %>% dplyr::filter(YEAR == 2012 | YEAR == 2013 | YEAR == 2014 |
                                YEAR == 2015 | YEAR == 2016) %>%
  group_by(YEAR, UNIQUE_CARRIER_NAME) %>% summarise(PASSENGERS = sum(PASSENGERS)) %>%
  ungroup %>% arrange(YEAR, desc(PASSENGERS))
```

```
top5_ref <- list()
j=1
for(i in 2012:2016) {
  temp <- dplyr::filter(top5, YEAR == i)[1:5,2]
  top5_ref[[j]] <- temp
  j <- j + 1
}
```

```
top5_ref <- as.data.frame(top5_ref)
colnames(top5_ref) <- c("2012", "2013", "2014", "2015", "2016")
kable(top5_ref[,1:3])
```

2012	2013	2014
Southwest Airlines Co.	Southwest Airlines Co.	Southwest Airlines Co.
Delta Air Lines Inc.	Delta Air Lines Inc.	Delta Air Lines Inc.
United Air Lines Inc.	United Air Lines Inc.	American Airlines Inc.
American Airlines Inc.	American Airlines Inc.	United Air Lines Inc.
US Airways Inc.	US Airways Inc.	US Airways Inc.

```
kable(top5_ref[,4:5])
```

2015	2016
Southwest Airlines Co.	Southwest Airlines Co.
Delta Air Lines Inc.	Delta Air Lines Inc.
American Airlines Inc.	American Airlines Inc.
United Air Lines Inc.	United Air Lines Inc.
JetBlue Airways	JetBlue Airways

It can be seen from the tables above that jet blue airways had made giant strides in securing a top 5 spot in 2015 and 2016 (since becoming classified as major airline in 2005) surpassing US Airways Inc.