DATA 607 Home Work 5

Dilip Ganesan

02/28/2017

Step 1: Loading the csv file and storing in data frame.

```
flighttimings=fread('flighttime.csv')
flighttimings=data.frame(flighttimings)
head(flighttimings)
```

```
##       V1      V2 LosAngeles Phoenix SanDiego SanFrancisco Seattle
## 1 ALASKA on time        497     221       212          503    1841
## 2        delayed         62      12        20          102     305
## 3                        NA      NA        NA           NA      NA
## 4 AMWEST on time        694    4840       383          320     201
## 5        delayed        117     415        65          129      61
```

```
knitr::kable(flighttimings)
```

V1	V2	LosAngeles	Phoenix	SanDiego	SanFrancisco	Seattle
ALASKA	on time	497	221	212	503	1841
	delayed	62	12	20	102	305
		NA	NA	NA	NA	NA
AMWEST	on time	694	4840	383	320	201
	delayed	117	415	65	129	61

Step 2: Cleaning of data.

```
# Removing the NA row from data frame.
flighttimings = flighttimings %>% na.omit()
# Adding Flight Names for Missing Rows
flighttimings[2,1]="ALASKA"
flighttimings[4,1]="AMWEST"

# Adding Column Header for the missing columns.
flighttimings=plyr::rename(flighttimings,c("V1"="Airline","V2"="Status"))
```

Step 3: Data Transformation using TIDYR

```
tidyflight = flighttimings %>%
    gather(cities, count, 3:7) %>%
    spread(Status, count)


knitr::kable(tidyflight)
```

Airline	cities	delayed	on time
ALASKA	LosAngeles	62	497
ALASKA	Phoenix	12	221
ALASKA	SanDiego	20	212
ALASKA	SanFrancisco	102	503

Airline	cities	delayed	on time
ALASKA	Seattle	305	1841
AMWEST	LosAngeles	117	694
AMWEST	Phoenix	415	4840
AMWEST	SanDiego	65	383
AMWEST	SanFrancisco	129	320
AMWEST	Seattle	61	201

Step 4: Data Analysis

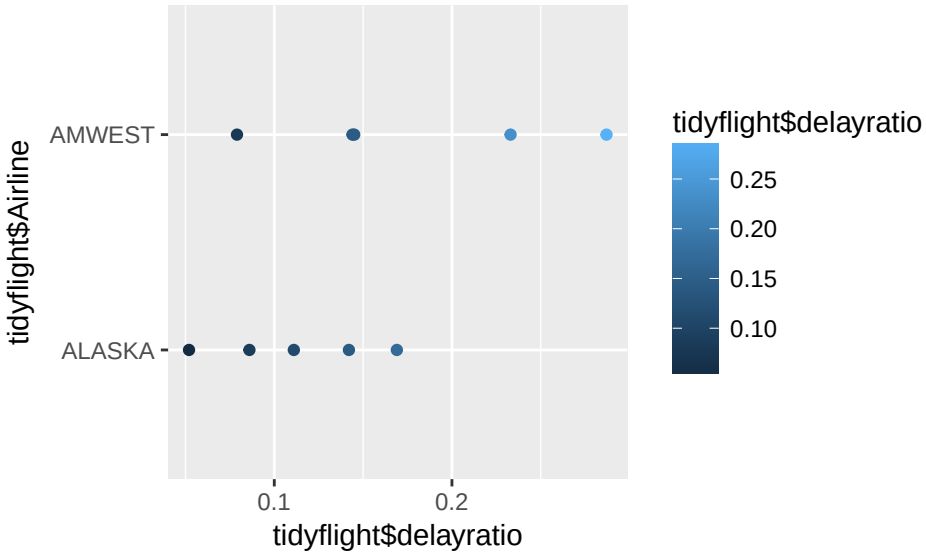
```
# Delay Calculation

tidyflight = tidyflight %>%
    mutate(total=`on time` + delayed, delayratio=round(delayed/total, 3))

knitr::kable(tidyflight)
```

Airline	cities	delayed	on time	total	delayratio
ALASKA	LosAngeles	62	497	559	0.111
ALASKA	Phoenix	12	221	233	0.052
ALASKA	SanDiego	20	212	232	0.086
ALASKA	SanFrancisco	102	503	605	0.169
ALASKA	Seattle	305	1841	2146	0.142
AMWEST	LosAngeles	117	694	811	0.144
AMWEST	Phoenix	415	4840	5255	0.079
AMWEST	SanDiego	65	383	448	0.145
AMWEST	SanFrancisco	129	320	449	0.287
AMWEST	Seattle	61	201	262	0.233

```
ggplot(tidyflight, aes(tidyflight$delayratio, tidyflight$Airline, colour = tidyflight$delayratio)) +
  geom_point()
```



Conclusion

The city delay ratios by airline show that ALASKA airlines have a lower individual delay ratio than AMWEST in every single city. But if the volume of flights are taken into account, then this might change.