Week 5 Assignment

Munkhnaran Gankhuyag

September 30, 2017

```
arrivals_delays <- read.csv("arrival_delays_data.csv", header = TRUE, sep = ",")
arrivals_delays
```

```
##         X     X.1 Los.Angeles Phoenix San.Diego San.Francisco Seattle
## 1  ALASKA on time         497     221       212           503    1841
## 2         delayed          62      12        20           102     305
## 3                          NA      NA        NA            NA      NA
## 4 AM WEST on time         694    4840       393           320     201
## 5         delayed         117     415        65           129      61
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
arrivals_delays2 <- slice(arrivals_delays,c(1,2,4,5))
arrivals_delays2
```

```
## # A tibble: 4 x 7
##         X     X.1 Los.Angeles Phoenix San.Diego San.Francisco Seattle
##    <fctr>  <fctr>       <int>   <int>     <int>         <int>   <int>
## 1  ALASKA on time         497     221       212           503    1841
## 2         delayed          62      12        20           102     305
## 3 AM WEST on time         694    4840       393           320     201
## 4         delayed         117     415        65           129      61
```

```
arrivals_delays2 <- data.frame(arrivals_delays2)
tidy <- arrivals_delays2 %>% gather(key = city, value = stat,3:7)
tidy <- data.frame(tidy)
tidy$stat <-as.numeric(tidy$stat)
tidy
```

```
##          X     X.1          city stat
## 1   ALASKA on time   Los.Angeles  497
## 2          delayed   Los.Angeles   62
## 3  AM WEST on time   Los.Angeles  694
## 4          delayed   Los.Angeles  117
## 5   ALASKA on time       Phoenix  221
## 6          delayed       Phoenix   12
## 7  AM WEST on time       Phoenix 4840
## 8          delayed       Phoenix  415
```

```
## 9   ALASKA on time     San.Diego  212
## 10         delayed     San.Diego   20
## 11 AM WEST on time     San.Diego  393
## 12         delayed     San.Diego   65
## 13  ALASKA on time San.Francisco  503
## 14         delayed San.Francisco  102
## 15 AM WEST on time San.Francisco  320
## 16         delayed San.Francisco  129
## 17  ALASKA on time       Seattle 1841
## 18         delayed       Seattle  305
## 19 AM WEST on time       Seattle  201
## 20         delayed       Seattle   61
```

```
colnames(tidy) <- c("airline","status","city","count")
tidy$airline <- c("ALASKA","ALASKA","AM WEST","AM WEST","ALASKA","ALASKA","AM WEST","AM WEST","ALASKA",
tidy
```

```
##    airline  status          city count
## 1   ALASKA on time   Los.Angeles   497
## 2   ALASKA delayed   Los.Angeles    62
## 3  AM WEST on time   Los.Angeles   694
## 4  AM WEST delayed   Los.Angeles   117
## 5   ALASKA on time       Phoenix   221
## 6   ALASKA delayed       Phoenix    12
## 7  AM WEST on time       Phoenix  4840
## 8  AM WEST delayed       Phoenix   415
## 9   ALASKA on time     San.Diego   212
## 10  ALASKA delayed     San.Diego    20
## 11 AM WEST on time     San.Diego   393
## 12 AM WEST delayed     San.Diego    65
## 13  ALASKA on time San.Francisco   503
## 14  ALASKA delayed San.Francisco   102
## 15 AM WEST on time San.Francisco   320
## 16 AM WEST delayed San.Francisco   129
## 17  ALASKA on time       Seattle  1841
## 18  ALASKA delayed       Seattle   305
## 19 AM WEST on time       Seattle   201
## 20 AM WEST delayed       Seattle    61
```

```
tidy_sort_by_city <-
  tidy %>%
  select(city,everything())
tidy_sort_by_city
```

```
##            city airline  status count
## 1    Los.Angeles  ALASKA on time   497
## 2    Los.Angeles  ALASKA delayed    62
## 3    Los.Angeles AM WEST on time   694
## 4    Los.Angeles AM WEST delayed   117
## 5        Phoenix  ALASKA on time   221
## 6        Phoenix  ALASKA delayed    12
## 7        Phoenix AM WEST on time  4840
## 8        Phoenix AM WEST delayed   415
## 9      San.Diego  ALASKA on time   212
## 10     San.Diego  ALASKA delayed    20
```

```
## 11     San.Diego AM WEST on time   393
## 12     San.Diego AM WEST delayed    65
## 13 San.Francisco  ALASKA on time   503
## 14 San.Francisco  ALASKA delayed   102
## 15 San.Francisco AM WEST on time   320
## 16 San.Francisco AM WEST delayed   129
## 17       Seattle  ALASKA on time  1841
## 18       Seattle  ALASKA delayed   305
## 19       Seattle AM WEST on time   201
## 20       Seattle AM WEST delayed    61
```

```
On_time_alaska <- filter(tidy_sort_by_city,status=='on time' & airline == 'ALASKA')
On_time_am_west <- filter(tidy_sort_by_city,status=='on time' & airline == 'AM WEST')
delayed_alaska <- filter(tidy_sort_by_city,status=='delayed' & airline == 'ALASKA')
delayed_am_west <- filter(tidy_sort_by_city,status=='delayed' & airline == 'AM WEST')

On_time_alaska <-On_time_alaska[ ,c(1,4)]
On_time_am_west <- On_time_am_west[ ,c(1,4)]
delayed_alaska <- delayed_alaska[ ,c(1,4)]
delayed_am_west <- delayed_am_west[ ,c(1,4)]
```

```
colnames(On_time_alaska) <- c("city","ALASKA")
colnames(On_time_am_west) <- c("city","AM West")
colnames(delayed_alaska) <- c("city","ALASKA")
colnames(delayed_am_west) <- c("city","AM West")
```

```
Alaska <- On_time_alaska
Alaska$Delayed <- delayed_alaska$ALASKA
Alaska$ALASKA <- as.numeric(Alaska$ALASKA)
Alaska$Delayed <- as.numeric(Alaska$Delayed)
Alaska$Total <- rowSums(Alaska[,2:3])
Alaska$Percent_Delayed_Alaska <- round(((Alaska$Delayed/Alaska$Total)*100), 2)
Alaska
```

```
##            city ALASKA Delayed Total Percent_Delayed_Alaska
## 1   Los.Angeles    497      62   559                  11.09
## 2       Phoenix    221      12   233                   5.15
## 3     San.Diego    212      20   232                   8.62
## 4 San.Francisco    503     102   605                  16.86
## 5       Seattle   1841     305  2146                  14.21
```

```
AM_West <- On_time_am_west
AM_West$Delayed <- delayed_am_west$`AM West`
AM_West$`AM West` <- as.numeric(AM_West$`AM West`)
AM_West$Delayed <- as.numeric(AM_West$Delayed)
AM_West$Total <- rowSums(AM_West[,2:3])
AM_West$Percent_Delayed_AM_West <- round(((AM_West$Delayed/AM_West$Total)*100), 2)
AM_West
```

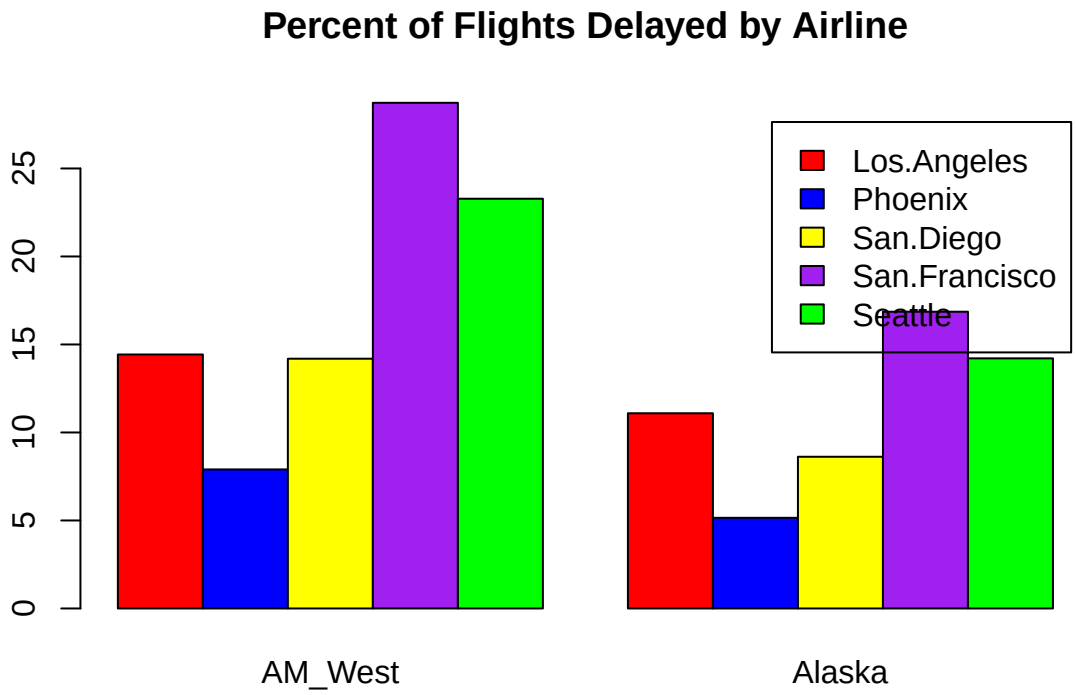
```
##            city AM West Delayed Total Percent_Delayed_AM_West
## 1   Los.Angeles     694     117   811                   14.43
## 2       Phoenix    4840     415  5255                    7.90
## 3     San.Diego     393      65   458                   14.19
## 4 San.Francisco     320     129   449                   28.73
## 5       Seattle     201      61   262                   23.28
```

```r
Percent_Delayed <- matrix(c(AM_West$Percent_Delayed_AM_West,Alaska$Percent_Delayed_Alaska), ncol = 2)
rownames(Percent_Delayed) <- AM_West$city
colnames(Percent_Delayed) <- c("AM_West", "Alaska")

Percent_Delayed
```

```
##               AM_West Alaska
## Los.Angeles     14.43  11.09
## Phoenix          7.90   5.15
## San.Diego       14.19   8.62
## San.Francisco   28.73  16.86
## Seattle         23.28  14.21
```

```r
barplot(Percent_Delayed, main = 'Percent of Flights Delayed by Airline',beside = T, col=c("red", "blue"
        legend= rownames(Percent_Delayed))
```



From the data we can see that AM West has a higher percentage of delays in every city.