

Week 5 Assignment

```
library(tidyr)
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
```

```
flightInfo <- read.csv("https://raw.githubusercontent.com/FlowControl/Data607/master/Week5_Assignment/a
```

Gather the data, and transform it into the “long” format

```
flightInfo2 <- gather(flightInfo, airport, count, Los_Angeles:Seattle, factor_key=TRUE)
```

Group by the airline and status, and get some stats on the average number of flights, and total flights

```
flightInfo2 %>% group_by(airline, status) %>% summarise(mean=mean(count), sum=sum(count))
```

```
## Source: local data frame [4 x 4]
## Groups: airline [?]
##
##   airline status   mean   sum
##   <fctr> <fctr> <dbl> <int>
## 1 ALASKA delayed  100.2   501
## 2 ALASKA on_time  654.8  3274
## 3 AM_WEST delayed  157.4   787
## 4 AM_WEST on_time 1287.6  6438
```

Filter the data to get delays only, and then gather statistics of the average number of delays per airline, and the total number of delays per airline. It appears that AM_WEST has the most total delays of the two airports.

```
delayOnly <- filter(flightInfo2, status %in% "delayed")
```

```
delayOnly %>% group_by(airline, status) %>% summarise(mean=mean(count), sum=sum(count))
```

```
## Source: local data frame [2 x 4]
## Groups: airline [?]
##
##   airline status   mean   sum
##   <fctr> <fctr> <dbl> <int>
## 1 ALASKA delayed  100.2   501
## 2 AM_WEST delayed  157.4   787
```