Airline Delays in the First Course (part 2: processing the data)

Nicholas Horton, nhorton@amherst.edu, January 17, 2014

Accessing the data

```
require(mosaic)
```

```
# reads in ds2 from part 1
load(url("http://www.amherst.edu/~nhorton/airlines/GRB.Rd"))
ds2 = na.omit(ds2)  # get rid of missing values
alleagle = subset(ds2, UniqueCarrier == "OO")
allmesa = subset(ds2, UniqueCarrier == "MQ")
```

Define a rule (then test it)

```
compareI = function(airlinea, airlineb) {
    # difference in means greater than 30 minutes and both
standard deviations <
    # 60
    meana = mean(airlinea)
    meanb = mean(airlineb)
    diffmeans = meana - meanb
    sda = mean(airlinea)
    sdb = sd(airlineb)
    if ((max(c(sda, sdb) >= 60)) | (abs(diffmeans) < 30)) {
        return("NEITHER")
    } else return(ifelse(diffmeans < 0, "Airline A", "Airline
B"))
}
compareI(c(10, 10), c(-40, -40))
```

```
## [1] "Airline B"
```

```
compareI(c(10, 10), c(-10, -10))
```

```
## [1] "NEITHER"
```

```
compareI(c(10, 10), c(-10, 1000))
```

```
## [1] "NEITHER"
```

```
# observed result
americaneagle = c(-10, -9, -2, -1, 9, 13, 17, 54, 98, 236)
mesa = c(-22, -16, -14, -8, -5, 0, 0, 3, 4, 28)
mean(americaneagle)
```

```
## [1] 40.5
```

```
sd(americaneagle)
```

```
## [1] 76.4
```

```
mean(mesa)
```

```
## [1] -3
```

```
sd(mesa)
```

```
## [1] 13.92
```

```
compareI(americaneagle, mesa)
```

```
## [1] "Airline B"
```

```
compareI(sample(alleagle$ArrDelay, 10),
sample(allmesa$ArrDelay, 10))
```

```
## [1] "NEITHER"
```

```
compareI(sample(alleagle$ArrDelay, 10),
sample(allmesa$ArrDelay, 10))
```

```
## [1] "NEITHER"
```

```
res = do(2000) * compareI(sample(alleagle$ArrDelay, 10),
sample(allmesa$ArrDelay,
    10))
tally(~result, data = res)
```

```
##
## Airline A Airline B   NEITHER     Total
##         6       192      1802      2000
```

Define another rule (then test it)

```
compareII = function(airlinea, airlineb) {
    # difference in means greater than 30 minutes and 10% more
delays
    meana = mean(airlinea)
    meanb = mean(airlineb)
    diffmeans = meana - meanb
    propdelays = mean(airlinea > 15) - mean(airlineb > 15)
    if (diffmeans > 30 & propdelays >= 0.1) {
        return("Airline B")
    } else if (diffmeans < -30 & propdelays <= -0.1) {
        return("Airline A")
    } else return("NEITHER")
}
compareII(c(0, 0, 0, 0, 0, 0, 20, 20, 20, 20), c(0, 0, 0, 0,
200, 200, 200,
    200, 200, 200))
```

```
## [1] "Airline A"
```

```
compareII(c(0, 0, 0, 0, 200, 200, 200, 200, 200, 200), c(0, 0,
0, 0, 0, 0, 20,
    20, 20, 20))
```

```
## [1] "Airline B"
```

```
compareII(c(0, 0, 0, 0, 200, 200, 200, 200, 200, 200), c(0, 0,
0, 20, 20, 20,
    20, 20, 20, 20))
```

```
## [1] "NEITHER"
```

```
compareII(americaneagle, mesa)
```

```
## [1] "Airline B"
```

```
res = do(2000) * compareII(sample(alleagle$ArrDelay, 10),
sample(allmesa$ArrDelay,
    10))
tally(~result, data = res)
```

```
##
## Airline A Airline B   NEITHER     Total
##        34       181      1785      2000
```