ITC pt4

Fayre-Ella Ooi

2025-04-14

```
# data set
CircuitOutage = read.csv(file = "/Users/fayreooi/Desktop/circuitWRegions.csv")
LookUp = read.csv(file = "/Users/fayreooi/Downloads/LookUpUpdate.csv")


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
```

```
df = left_join(CircuitOutage, LookUp, by = "Circuit.Name")

# clean

# make values numerical
df$Outage.Duration..min. = as.numeric(gsub(",", "", df$Outage.Duration..min.))
df$Customers.Affected = as.numeric(gsub(",", "", df$Customers.Affected))
df$Customer.Count = as.numeric(gsub(",", "", df$Customer.Count))

LookUp$Customer.Count = as.numeric(gsub(",", "", LookUp$Customer.Count))


# drop unnecessary columns
df = df[, !(names(df) %in% c("Region.x", "X", "Circuit.Number.y"))]

# add new column
df$Percentage.Customers.Affected = df$Customers.Affected / df$Customer.Count

#
```

```
# find the mean outage duration per outage cause
sort(tapply(df$Outage.Duration..min., df$Outage.Cause, mean))
```

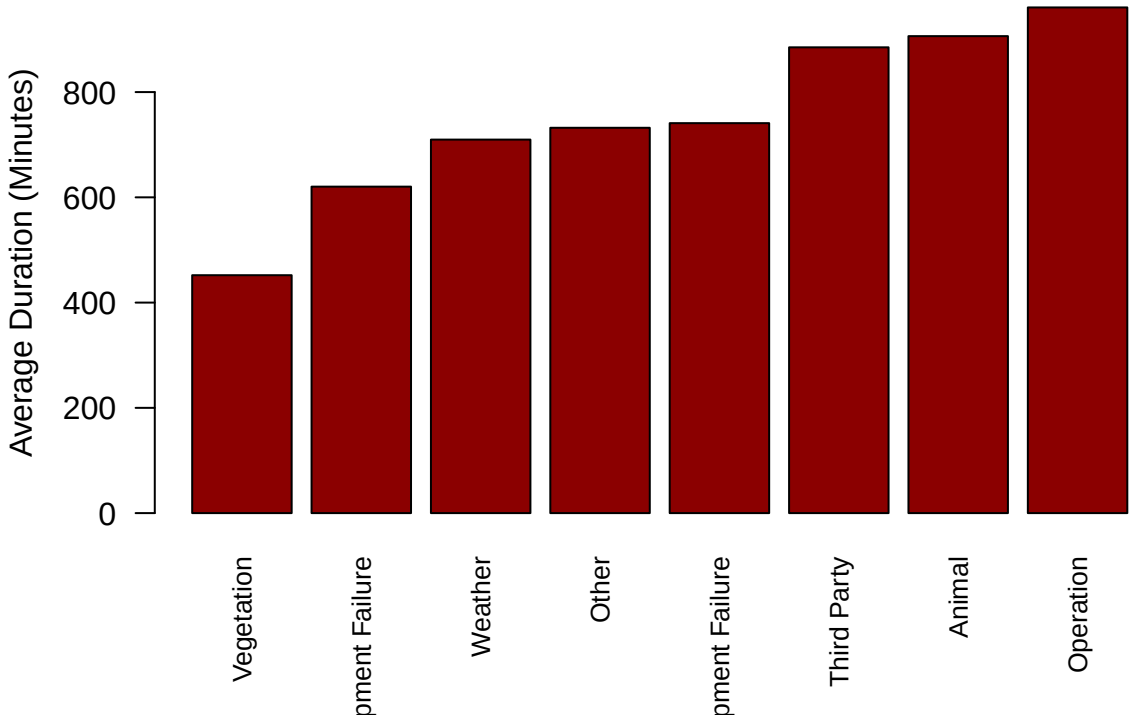
```
##           Vegetation OH Equipment Failure              Weather
##               452.0000             620.2941             709.6667
##                Other UG Equipment Failure          Third Party
##               732.0909             740.9167             885.0000
```

```
##               Animal            Operation
##             906.2857             960.8000
```

```r
# - operation has the longest mean outage duration

# find the amount of outages per outage cause
sort(tapply(df$Outage.Duration..min., df$Outage.Cause, length))
```

```
##           Vegetation               Animal            Operation
##                    5                    7                   10
##                Other              Weather          Third Party
##                   11                   12                   14
## OH Equipment Failure UG Equipment Failure
##                   17                   24
```

```r
# UG equipment failure has the most amount of outages

# make dataframe of outage causes and average outage duration
df1 = matrix(c(452, 620.2941, 709.6667, 732.0909, 740.9167, 885, 906.2857,
               960.8),
             nrow = 1, ncol = 8,
             dimnames = list(rownames = c("Average Outage Duration"),
                             colnames = c("Vegetation", "OH Equipment Failure",
                                          "Weather", "Other", "UG Equipment Failure",
                                          "Third Party", "Animal", "Operation")))

# plot data
barplot(df1,
        main = "Average Outage Duration by Cause",
        ylab = "Average Duration (Minutes)",
        las = 2,
        col = "darkred",
        cex.names = 0.8)
```




```
# perform chisquare test
X2.OD.cause = chisq.test(df1)
X2.OD.cause$expected
```

```
## [1] 750.8818 750.8818 750.8818 750.8818 750.8818 750.8818 750.8818 750.8818
```

```
X2.OD.cause$residuals
```

```
## [1] -10.9072091  -4.7655866  -1.5040774  -0.6857423  -0.3636589   4.8944293
## [7]   5.6712167   7.6606283
```

```
X2.OD.cause$p.value
```

```
## [1] 2.835572e-52
```

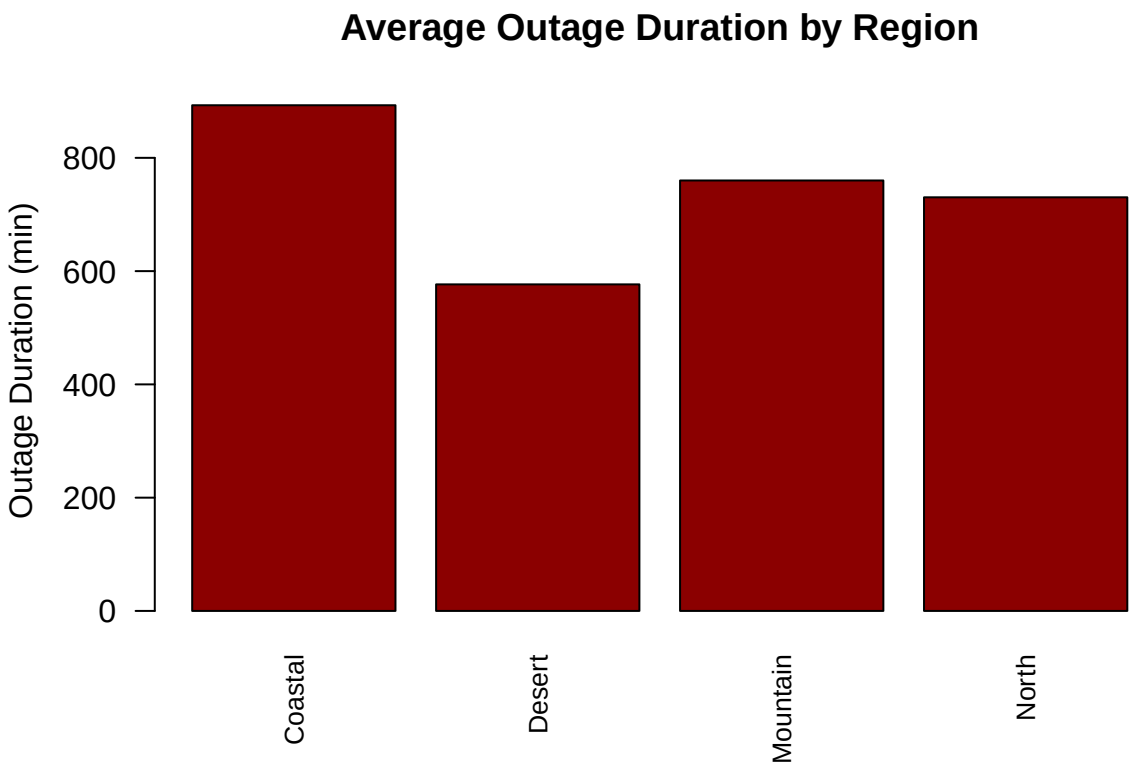
```
# causes matter
# specifically operation, animal, and third party
```

```
# outage duration per region
tapply(df$Outage.Duration..min., df$Region.y, mean)
```

```
##  Coastal   Desert Mountain    North
## 892.8571 576.5556 760.0000 730.3226
```

```
# make dataframe
df2 = matrix(c(892.8571, 576.5556, 760, 730.3226),
             ncol = 4, nrow = 1,
             dimnames = list(rownames = c("Average Outage Duration"),
                             colnames = c("Coastal", "Desert", "Mountain", "North")))

# barplot of data
barplot(df2,
```

```
        main = "Average Outage Duration by Region",
        ylab = "Outage Duration (min)",
        las = 2,
        col = "darkred",
        cex.names = 0.8)
```




```
# perform chisquare test
X2.OD.regions = chisq.test(df2)
X2.OD.regions$expected
```

```
## [1] 739.9338 739.9338 739.9338 739.9338
```

```
X2.OD.regions$residuals
```

```
## [1]  5.6218228 -6.0061716  0.7376802 -0.3533315
```

```
X2.OD.regions$p.value
```

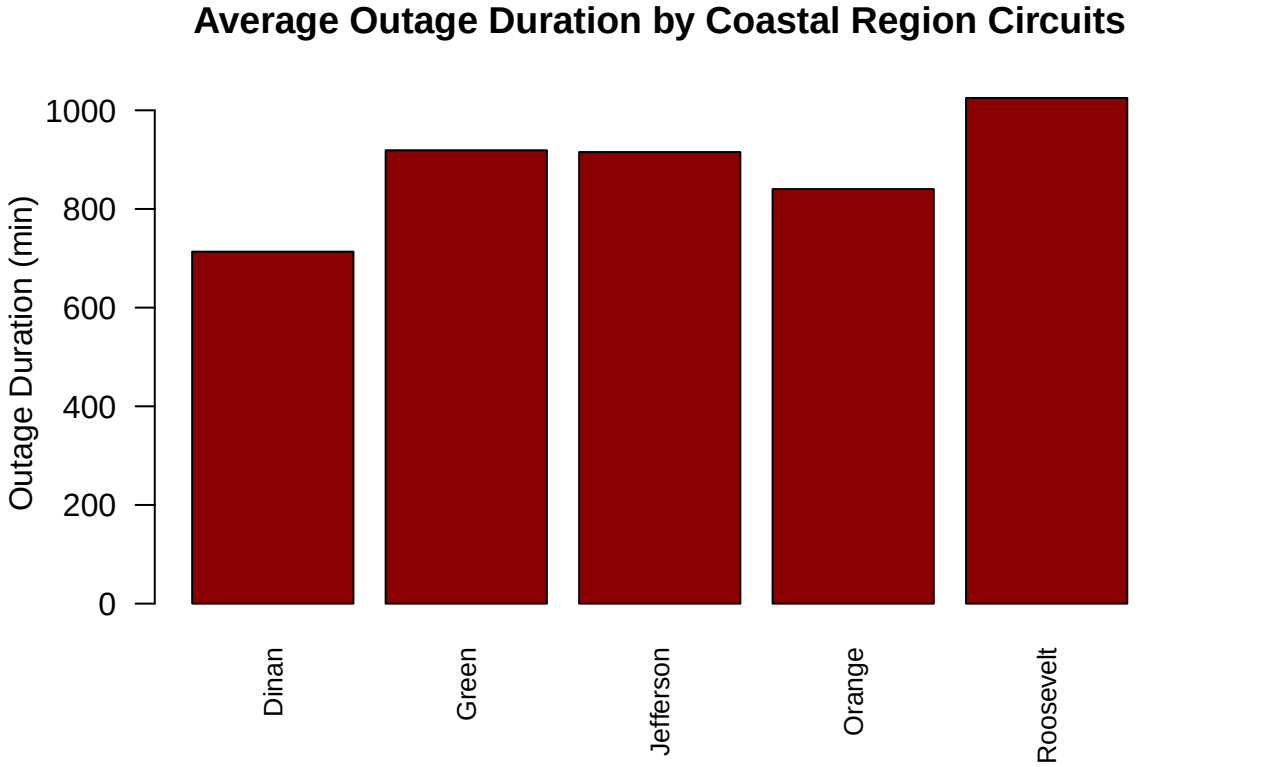
```
## [1] 9.637023e-15
```

```
# FOCUS: COASTAL
temp = subset(df, Region.y %in% c("Coastal"))

tapply(temp$Outage.Duration..min., temp$Circuit.Name, mean)
```

```
##     Dinan     Green Jefferson    Orange Roosevelt
##  713.4286  919.0000  915.3333  840.3333 1025.0000
```

```
df.coastal = matrix(c(713.4286, 919, 915.3333, 840.3333, 1025),
                    ncol = 5, nrow = 1,
                    dimnames = list(rownames = c("Average Outage Duration"),
                                    colnames = c("Dinan", "Green", "Jefferson",
                                                 "Orange", "Roosevelt")))
```

```r
# barplot of data
barplot(df.coastal,
        main = "Average Outage Duration by Coastal Region Circuits",
        ylab = "Outage Duration (min)",
        las = 2,
        col = "darkred",
        cex.names = 0.8)
```




```r
# perform chisquare test
X2.OD.coastal = chisq.test(df.coastal)
X2.OD.coastal$expected
```

```
## [1] 882.619 882.619 882.619 882.619 882.619
```

```r
X2.OD.coastal$residuals
```

```
## [1] -5.694940  1.224581  1.101160 -1.423336  4.792535
```

```r
X2.OD.coastal$p.value
```

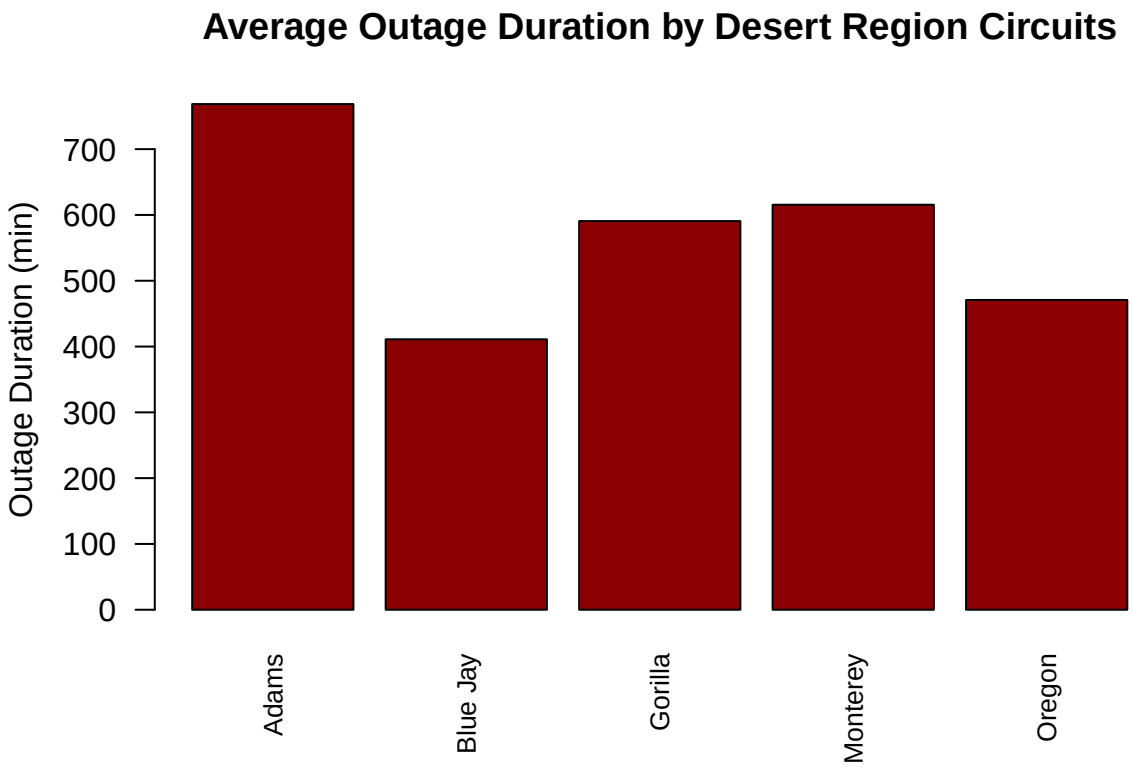
```
## [1] 2.712468e-12
```

```r
# FOCUS: DESERT
temp = subset(df, Region.y %in% c("Desert"))

tapply(temp$Outage.Duration..min., temp$Circuit.Name, mean)
```

```
##    Adams Blue Jay  Gorilla Monterey   Oregon
## 768.5000 411.0000 590.6667 615.5000 470.7500
```

```r
df.desert = matrix(c(768.5, 411, 590.6667, 615.5, 470.75),
                   ncol = 5, nrow = 1,
```

```
                    dimnames = list(rownames = c("Average Outage Duration"),
                                    colnames = c("Adams", "Blue Jay", "Gorilla",
                                                 "Monterey", "Oregon")))

# barplot of data
barplot(df.desert,
        main = "Average Outage Duration by Desert Region Circuits",
        ylab = "Outage Duration (min)",
        las = 2,
        col = "darkred",
        cex.names = 0.8)
```




```
# perform chisquare test
X2.OD.desert = chisq.test(df.desert)
X2.OD.desert$expected
```

```
## [1] 571.2833 571.2833 571.2833 571.2833 571.2833
```

```
X2.OD.desert$residuals
```

```
## [1]  8.2512134 -6.7059854  0.8109672  1.8499507 -4.2061459
```

```
X2.OD.desert$p.value
```

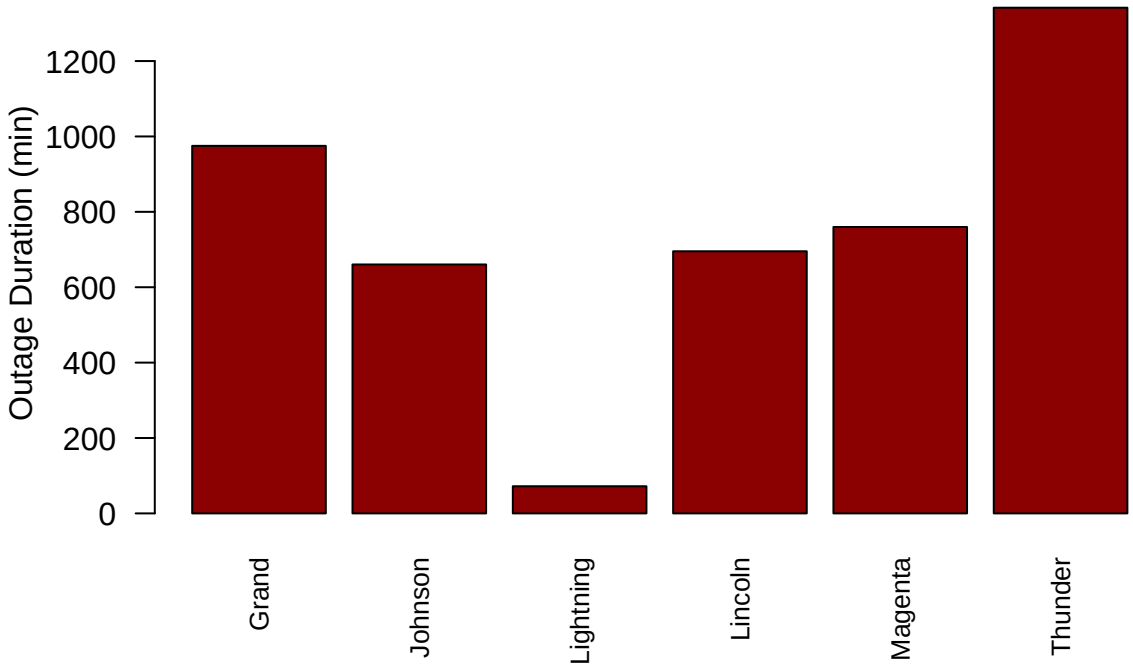
```
## [1] 3.617301e-28
```

```
# FOCUS: MOUNTAIN
temp = subset(df, Region.y %in% c("Mountain"))

tapply(temp$Outage.Duration..min., temp$Circuit.Name, mean)
```

```
##     Grand   Johnson Lightning   Lincoln   Magenta   Thunder
```

```
##  975.0000  660.3333   72.0000  695.3333  759.7000 1341.5000
```

```
df.mountain = matrix(c(975, 660.3333, 72, 695.3333, 759.7, 1341.5),
                     ncol = 6, nrow = 1,
                     dimnames = list(rownames = c("Average Outage Duration"),
                                     colnames = c("Grand", "Johnson", "Lightning",
                                                  "Lincoln", "Magenta", "Thunder")))

# barplot of data
barplot(df.mountain,
        main = "Average Outage Duration by Mountain Region Circuits",
        ylab = "Outage Duration (min)",
        las = 2,
        col = "darkred",
        cex.names = 0.8)
```




```
# perform chisquare test
X2.OD.mountain = chisq.test(df.mountain)
X2.OD.mountain$expected
```

```
## [1] 750.6444 750.6444 750.6444 750.6444 750.6444 750.6444
```

```
X2.OD.mountain$residuals
```

```
## [1]   8.1887897  -3.2962805 -24.7699515  -2.0188099   0.3305206  21.5657317
```

```
X2.OD.mountain$p.value
```

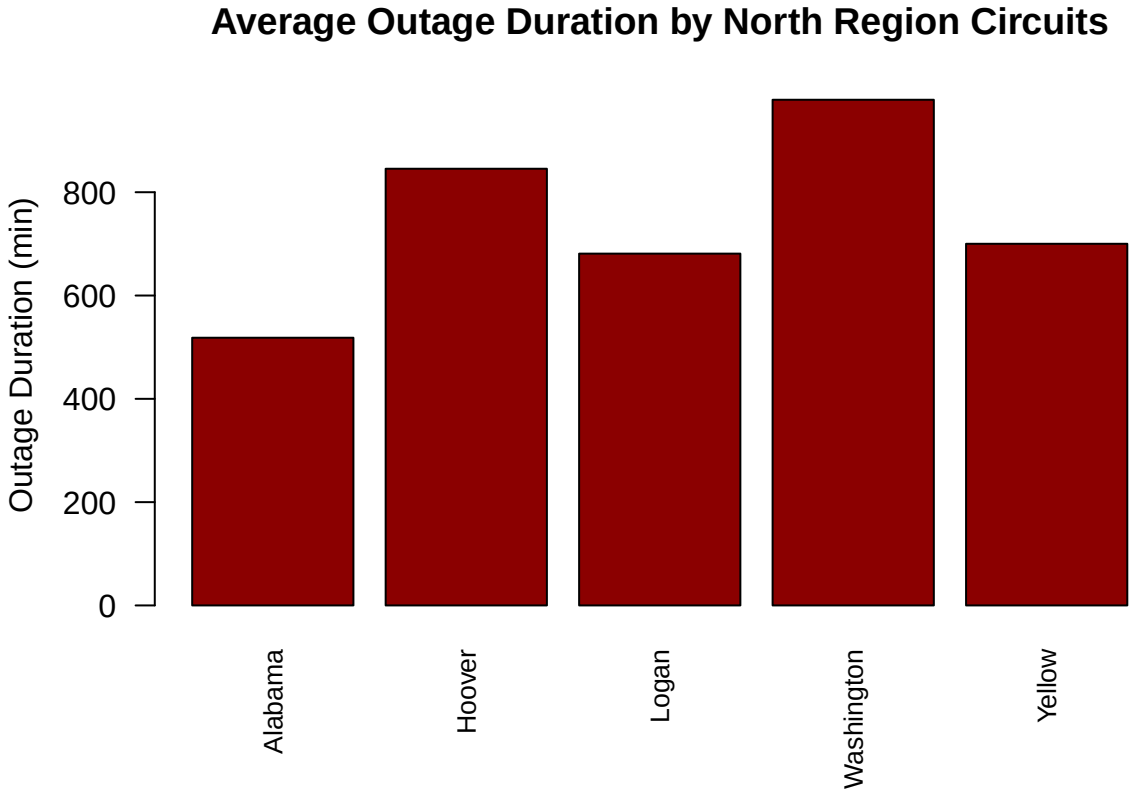
```
## [1] 9.376012e-249
```

```
# FOCUS: NORTH
temp = subset(df, Region.y %in% c("North"))
```

```
tapply(temp$Outage.Duration..min., temp$Circuit.Name, mean)
```

```
##     Alabama      Hoover      Logan Washington      Yellow
##   518.2500   845.4286   681.1000   979.0000   700.1429
```

```
df.north = matrix(c(518.25, 845.4286, 681.1, 979, 700.1429),
                  ncol = 5, nrow = 1,
                  dimnames = list(rownames = c("Average Outage Duration"),
                                  colnames = c("Alabama", "Hoover", "Logan",
                                               "Washington", "Yellow")))

# barplot of data
barplot(df.north,
        main = "Average Outage Duration by North Region Circuits",
        ylab = "Outage Duration (min)",
        las = 2,
        col = "darkred",
        cex.names = 0.8)
```




```
# perform chisquare test
X2.OD.north = chisq.test(df.north)
X2.OD.north$expected
```

```
## [1] 744.7843 744.7843 744.7843 744.7843 744.7843
```

```
X2.OD.north$residuals
```

```
## [1] -8.300776  3.687856 -2.333550  8.582242 -1.635771
```

```
X2.OD.north$p.value
```

```
## [1] 1.766132e-34
```