Introduction to R Programming
Lecture 4

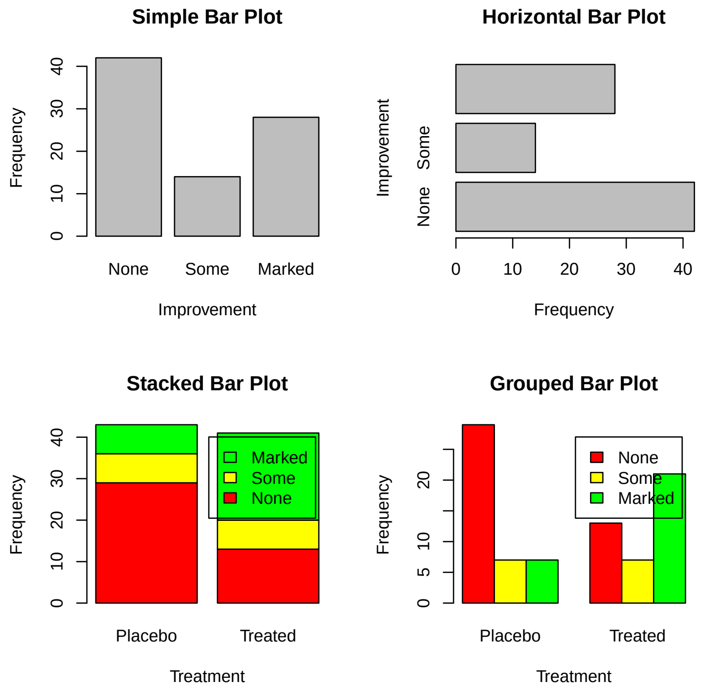
Minghao Wu
*minghaowu*_2015@163.*com*

April 16, 2015

1 Baisc Graph

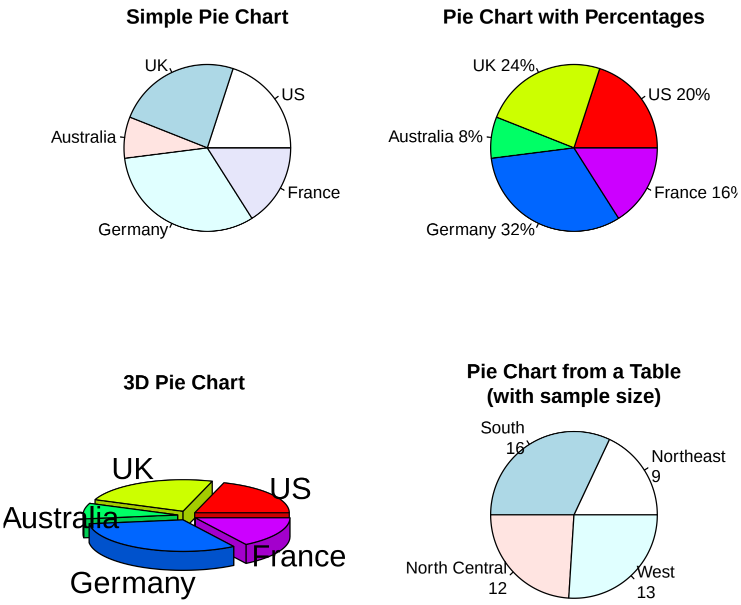
1.1 Bar Plot

```
> library(vcd)
> counts <- table(Arthritis$Improved)
> counts

  None   Some Marked
    42     14     28

> par(mfrow=c(2,2))
> barplot(counts,
+         main="Simple Bar Plot",
+         xlab="Improvement", ylab="Frequency")
> barplot(counts,
+         main="Horizontal Bar Plot",
+         xlab="Frequency", ylab="Improvement",
+         horiz=TRUE)
> counts <- table(Arthritis$Improved, Arthritis$Treatment)
> counts

         Placebo Treated
  None        29      13
  Some         7       7
  Marked       7      21

> barplot(counts,
+         main="Stacked Bar Plot",
+         xlab="Treatment", ylab="Frequency",
+         col=c("red", "yellow","green"),
+         legend=rownames(counts))
> barplot(counts,
+    main="Grouped Bar Plot",
+    xlab="Treatment", ylab="Frequency",
+    col=c("red", "yellow", "green"),
+    legend=rownames(counts), beside=TRUE)
>
```



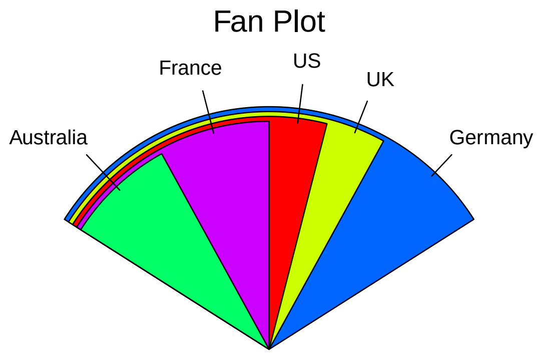
1.2 Pie Chart

```
> library(plotrix)
> par(mfrow=c(2,2))
> slices <- c(10, 12,4, 16, 8)
> lbls <- c("US", "UK", "Australia", "Germany", "France")
> pie(slices, labels = lbls,main="Simple Pie Chart",edges=300,radius=1)
> pct <- round(slices/sum(slices)*100)
> lbls2 <- paste(lbls, " ", pct, "%", sep="")
> pie(slices, labels=lbls2, col=rainbow(length(lbls2)),
+  main="Pie Chart with Percentages",edges=300,radius=1)
> pie3D(slices, labels=lbls,explode=0.1,
+   main="3D Pie Chart ",edges=300,radius=1)
> mytable <- table(state.region)
> lbls3 <- paste(names(mytable), "\n", mytable, sep="")
> pie(mytable,labels=lbls3,
+     main="Pie Chart from a Table\n(with sample size)",
+     edges=300,radius=1)
> slices <- c(10, 12,4, 16, 8)
> lbls <- c("US", "UK", "Australia", "Germany", "France")
>
```



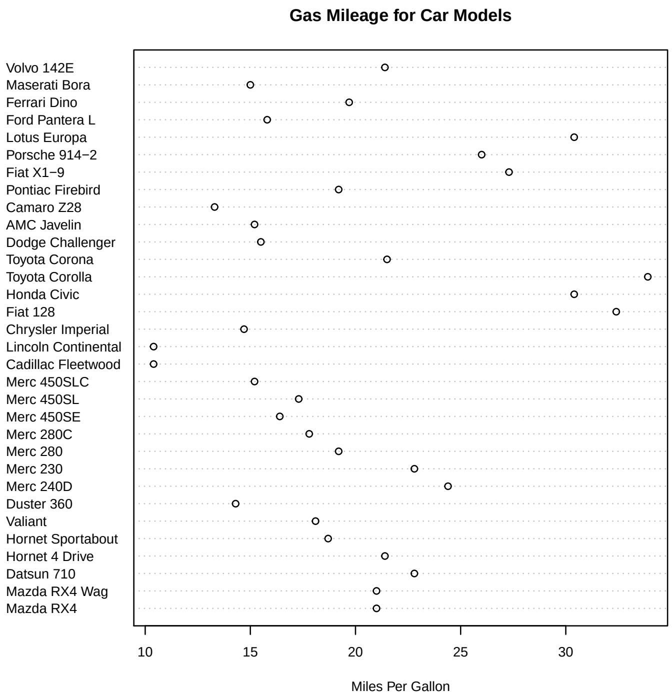
1.3 Fan Plot

```
> fan.plot(slices, labels = lbls, main="Fan Plot")
```



1.4 Dot Plot

```
> dotchart(mtcars$mpg,
+          labels=row.names(mtcars),cex=0.7,
+          main="Gas Mileage for Car Models",
+          xlab="Miles Per Gallon")
```



2 Basic Statistics

2.1 Descriptive statistics

```
> head(mtcars)
```

```
                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
> summary(mtcars)
```

```
      mpg             cyl             disp             hp       
 Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
 1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
 Median :19.20   Median :6.000   Median :196.3   Median :123.0  
 Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
 3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
 Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
      drat             wt             qsec             vs        
 Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
 1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
 Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
 Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
 3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
 Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
       am              gear            carb      
 Min.   :0.0000   Min.   :3.000   Min.   :1.000  
 1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
 Median :0.0000   Median :4.000   Median :2.000  
 Mean   :0.4062   Mean   :3.688   Mean   :2.812  
 3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
 Max.   :1.0000   Max.   :5.000   Max.   :8.000  
```

2.2 Frequency and contingency tables

```
> attach(mtcars)
> table(cyl)

cyl
 4  6  8 
11  7 14 

> summary(mpg)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  10.40   15.42   19.20   20.09   22.80   33.90 

> table(cut(mpg,seq(10,34,by=2)))

(10,12] (12,14] (14,16] (16,18] (18,20] (20,22] (22,24] (24,26] (26,28] (28,30] 
      2       1       7       3       5       5       2       2       1       0 
(30,32] (32,34] 
      2       2 
```

2.3 Correlations

```
> states = state.x77[,1:6]
> cov(states)

              Population      Income   Illiteracy     Life Exp      Murder
Population 19931683.7588 571229.7796  292.8679592 -407.8424612 5663.523714
Income       571229.7796 377573.3061 -163.7020408  280.6631837 -521.894286
Illiteracy      292.8680   -163.7020    0.3715306   -0.4815122    1.581776
Life Exp       -407.8425    280.6632   -0.4815122    1.8020204   -3.869480
```

```
Murder         5663.5237   -521.8943    1.5817755   -3.8694804   13.627465
HS Grad       -3551.5096   3076.7690   -3.2354694    6.3126849  -14.549616
                HS Grad
Population -3551.509551
Income      3076.768980
Illiteracy    -3.235469
Life Exp       6.312685
Murder       -14.549616
HS Grad       65.237894
```

```
> var(states)

              Population      Income   Illiteracy     Life Exp      Murder
Population 19931683.7588 571229.7796  292.8679592 -407.8424612 5663.523714
Income       571229.7796 377573.3061 -163.7020408  280.6631837 -521.894286
Illiteracy      292.8680   -163.7020    0.3715306   -0.4815122    1.581776
Life Exp       -407.8425    280.6632   -0.4815122    1.8020204   -3.869480
Murder         5663.5237   -521.8943    1.5817755   -3.8694804   13.627465
HS Grad       -3551.5096   3076.7690   -3.2354694    6.3126849  -14.549616
                HS Grad
Population -3551.509551
Income      3076.768980
Illiteracy    -3.235469
Life Exp       6.312685
Murder       -14.549616
HS Grad       65.237894
```

```
> cor(states)

            Population     Income Illiteracy    Life Exp     Murder     HS Grad
Population  1.00000000  0.2082276  0.1076224 -0.06805195  0.3436428 -0.09848975
Income      0.20822756  1.0000000 -0.4370752  0.34025534 -0.2300776  0.61993232
Illiteracy  0.10762237 -0.4370752  1.0000000 -0.58847793  0.7029752 -0.65718861
Life Exp   -0.06805195  0.3402553 -0.5884779  1.00000000 -0.7808458  0.58221620
Murder      0.34364275 -0.2300776  0.7029752 -0.78084575  1.0000000 -0.48797102
HS Grad    -0.09848975  0.6199323 -0.6571886  0.58221620 -0.4879710  1.00000000
```

2.4 T-test

```
> x = rnorm(100, mean = 10, sd = 1)
> y = rnorm(100, mean = 30, sd = 10)
> t.test(x, y, alt = "two.sided",paired=TRUE)

        Paired t-test

data:  x and y
t = -20.8901, df = 99, p-value < 2.2e-16
```

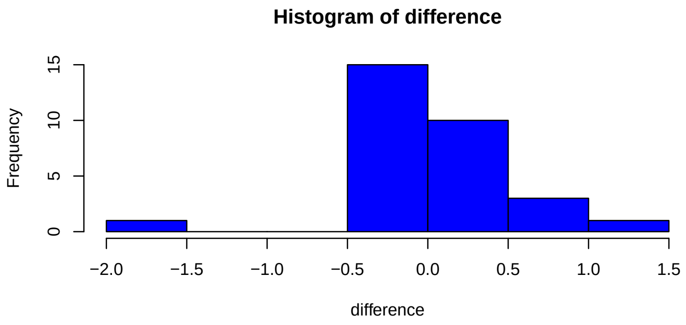
```
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -21.35301 -17.64851
sample estimates:
mean of the differences
              -19.50076
```

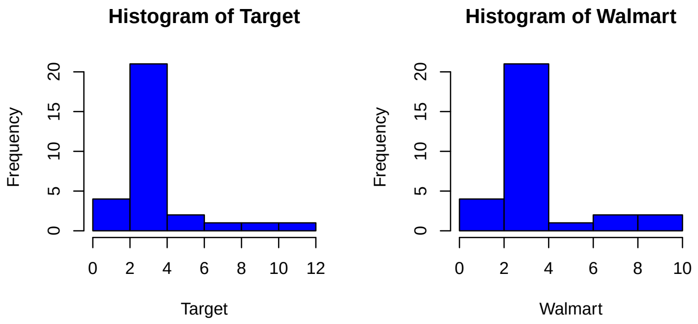
2.5 Nonparametric tests of group differences

```
> wilcox.test(x,y,alt="less")

        Wilcoxon rank sum test with continuity correction

data:  x and y
W = 61, p-value < 2.2e-16
alternative hypothesis: true location shift is less than 0
```

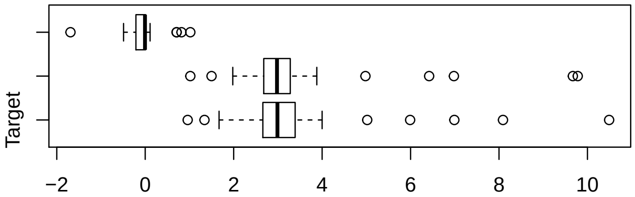
3 Practical Example

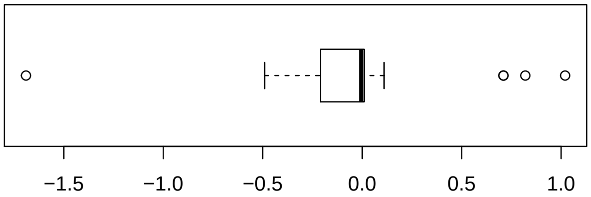
```
> data=read.csv("~/documents/R Programming/project/STAT.csv")
> attach(data)
> layout(matrix(c(1,1,2,3),2,2,byrow=TRUE))
> hist(difference,col="blue")
> hist(Target,col="blue")
> hist(Walmart,col="blue")
```





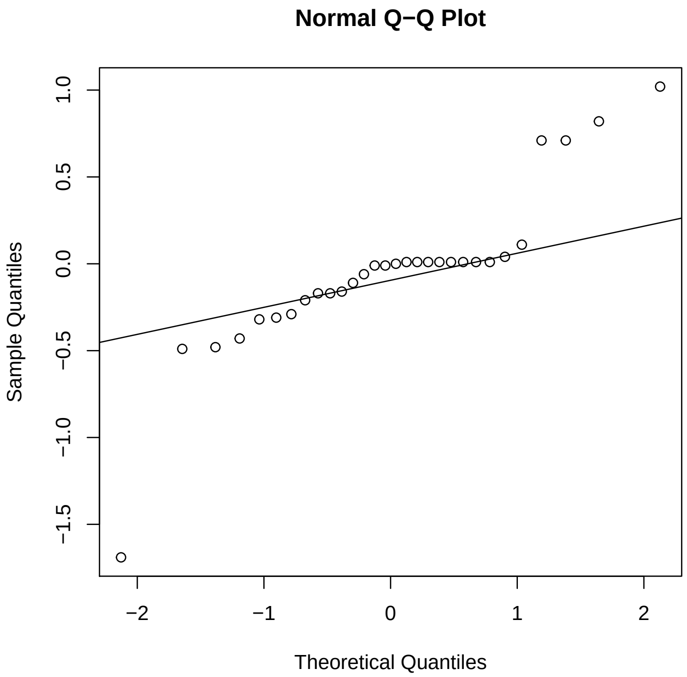
```
> par(mfrow=c(2,1))
> boxplot(data[2:4],horizontal=TRUE)
> boxplot(difference,horizontal=TRUE)
```





```
> par(mfrow=c(1,1))
> qqnorm(difference)
> qqline(difference)
> binom.test(length(difference[difference>=0]),
+            length(difference),
+            alter="two.sided")

        Exact binomial test

data:  length(difference[difference >= 0]) and length(difference)
number of successes = 15, number of trials = 30, p-value = 1
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.3129703 0.6870297
sample estimates:
probability of success
                   0.5

> wilcox.test(difference,alter="two.sided")

        Wilcoxon signed rank test with continuity correction

data:  difference
```

```
V = 174.5, p-value = 0.3557
alternative hypothesis: true location is not equal to 0

> t.test(difference,alter="two.sided")

        One Sample t-test

data:  difference
t = -0.5432, df = 29, p-value = 0.5911
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.2255485  0.1308819
sample estimates:
  mean of x
-0.04733333
```



```
> library(nortest)
> ad.test(difference)

        Anderson-Darling normality test

data:  difference
A = 2.1234, p-value = 1.552e-05
```