Cheatsheet for Exam 1

Math

Mean: $\overline{x} = \frac{1}{n}\sum x_i$
Quartiles: Let M = median
1. First quartile Q_1 is the median of the data $< M$
2. Third quartile Q_3 is the median of the data $> M$
Standard deviation: $s = \sqrt{\frac{1}{n-1}\sum(x_i - \overline{x})^2}$
Variance: s^2
Correlation: $r = \frac{1}{n-1}\sum\left(\frac{x_i-\overline{x}}{s_x}\right)\left(\frac{y_i-\overline{y}}{s_y}\right)$
Least-squares regression line: $\hat{y} = a + bx$

$$b = r\,\frac{s_y}{s_x} \qquad a = \overline{y} - b\overline{x}$$

Graphing in R

Pie chart:
```
pie(percentages, categories,
    main = "title")
```
Bar chart:
```
barplot(percentages,
        names.arg = categories,
        col = "color",
        xlab = "x-axis label",
        ylab = "y-axis label",
        main = "title")
```
Histogram:
```
hist(data, breaks = classes,
     col = "color",
     xlab = "x-axis label",
     main = "title")
```
Dotplot:
```
stripchart(data,
           method = "stack",
           offset = .5, at = 0,
           pch = 19,
           col = "color",
           xlab = "x-axis label",
           frame.plot = FALSE)
```
Boxplot:
```
boxplot(data1, data2,
        names = c("var_name1", "var_name2"),
        col = "color",
        ylab = "y-axis label",
        main = "title",
        outline = FALSE)
```
Scatterplot:
```
plot(x_data, y_data,
     xlab = "x-axis label",
     ylab = "y-axis label",
     main = "title",
     col = "color",
     pch = 19,
     cex = 2)
```
Add least-squares regression line to scatterplot:
```
abline(model,
       col = "orangered",
       lwd = 3)
```

Statistical commands in R

Formating output: `cat("Text", variable_name, "\n")`
Median: `median(data)`
Mean: `mean(data)`
Sorting data from lowest to largest value: `sort(data)`
Five-number summary: `fivenum(data)`
Standard deviation: `sd(data)`
Variance: `var(data)`
Correlation: `cor(x_data,y_data)`

Least-squares regression line:
```
model <- lm(y_data ~ x_data)
print(model)
```
Create a two-way table:
```
data_in_matrix_form <- matrix(data,
                              nrow = m, ncol = n)
rownames(data_in_matrix_form) <- row_classes
colnoames(data_in_matrix_form) <- column_classes
two_way_table <- as.table(data_in_matrix_form)
```
Print a two-way table: `print(two_way_table)`
The total number of observations in a two-way table:
```
margin.table(two_way_table)
```
Marginal distributions:
```
margin.table(two_way_table, margin = 1)
margin.table(two_way_table, margin = 2)
```
Conditional distributions:
```
prop.table(two_way_table, margin=1)
prop.table(two_way_table, margin=2)
```