Data Analysis Course Outline

WEEK 1

- Why use code!?
- INSTALL R / R-studio
- Familiarize R-studio functions and layout
- Where to look for help

```
# Functions covered (parital list):
help()
```

WEEK 2

- Assigning things to objects
- Get familiar with object types and basic functions
  - values, vectors, lists
  - data frames and matrices
  - boolean, character, numeric, POSIXct

- Accessing elements of objects
- Boolean evaluations
- Data-type conversions

```
# Functions covered (parital list):
=       <-      ->
class()
data.frame()  as.factor()   as.numeric()    as.character()    as.POSIXct()    as.matrix()
==    <   >   <=    >=
+     -   *   /
which()   signif()    ceiling()     floor()       round()
c()       list()      cbind()       rbind()       sum()         mean()        sd()
```

WEEK 3

- Importing data
- Useful data formats
- Data structure and attributes
- Summary stats and basic visualizations
- Exploring data
  - Sorting, Transposing, Sampling
  - heatmaps, boxplots, barcharts, scatterplots, histograms

```
# Functions covered (parital list):
read.csv()        read.delim()
str()             dim()             names()           attributes()          head()
summary()         min()             max()             range()               quantile()
hist()            boxplot()         barplot()         plot()                heatmap()
sample()          t()               sort()            tail()                var()
```

WEEK 4

- Finding/Installing/Loading packages
- Extending functionality of R
- Subsetting and manipulating raw data
- Output options

```
# Functions covered (parital list):
install.packages()
library()
cor()
write.table()       sink()        tiff/jpeg/png() / dev.off()
```

```
  Skills Test 1:

  * Import data set
  * Convert elements to new data type
  * Subset based on values
  * Calculate summary statistics
  * Create basic summary figures
  * Export summary statistics to text file
```

WEEK 5

- Data estimations
  - point estimates
  - interval estimates
- Hypothesis testing / Model fitting
  - t-test (paired/unpaired)
  - chi-square
  - ANOVA
  - LM/GLM

```
# Functions covered (parital list):
lm()        glm()         aov()       summary()
t.test()    chisq.test()
```

WEEK 6

Experimental design common designs and analysis options quantitative vs qualitative data Probability distributions Fitting distributions Type I and Type II errors Post-hoc tests

```
# Packages used (partial list):
fitdistrplus
MASS
# Functions covered (parital list):
plotdist()      descdist()
fitdist()       denscomp()        cdfcomp()
TukeyHSD()
```

WEEK 7

- Non-parametric alternatives

- Mann-Whitney-Wilcoxin
- Kruskal-Wallace
- Apply functions

```
# Packages used (partial list):

# Functions covered (parital list):
wilcox.test()     kruskal.test()

apply()     sapply()      lapply()      tapply()
```

WEEK 8

- Other peoples' data
- Principles of tidy data
- Intuitive manipulations and group functions
  - filter
  - arrange
  - select
  - mutate
  - group_by
  - summarize
  - %>%
- Tidy data transformations
  - gather
  - spread

```
# Packages used (partial list):
dplr      plyr      tidyr
# Functions covered (parital list):
filter()      arrange()       select()        mutate()
group_by()    summarize()     %>%
gather()      spread()
```

```
  Skills Test 2:

  * Import messy data
  * Convert to tidy format
  * Plot data distribution
  * Rearrange and mutate data set
  * Summary stats on grouped data
  * Test hypothesis / post-hoc tests
```

WEEK 9

- Predicting data
- Intro to ggplot

```
# Packages used (partial list):
ggplot2

# Functions covered (parital list):
predict()
```

```
qplot()
ggplot()
    aes()
```

WEEK 10

- Figure generation
- Figure export

```
# Packages used (partial list):
ggplot2

# Functions covered (parital list):
ggplot()
geom_point()      geom_boxplot()      geom_bar()      geom_violin()
labs()            ggsave()
```

WEEK 11

- Figure generation continued

```
# Packages used (partial list):
ggplot2
gridExtra

# Functions covered (parital list):
grid.arrange()
ggplot()
scale_*()
```

WEEK 12

- Data standardization / normalization
- Ecology examples
  - Ordinations / NMDS
  - PermANOVA
  - Distance measures
  - Diversity measures

```
# Packages used (partial list):
vegan

# Functions covered (parital list):
decostand()       rrarefy()         dist()        betadiver()
metaMDS()         adonis()          diversity()   betadisper()
```

```
  Skills Test 3:

  * Import data set
  * Fit appropriate model
  * Use model to predict new response values from new predictors
  * Generate and export plots from data sets
```

WEEK 13

- Importing and manipulating DNA sequence data
  - Bioconductor
  - Sequence data
  - Biostrings
- Phylogenetics examples
  - Sequence alignment
  - Tree building
  - Taxonomic assignment

```
# Packages used (partial list):
Bioconductor
ape
biostrings

# Functions covered (parital list):
```

WEEK 14

- Command-line tools
- BASH
- compression
- grep, sed, find, |

WEEK 15

- Data management
- Reporting
- Rmd

  Skills Test 4 (final):

  - Command-line data access and manipulation
  - Writing a script to
    - import specific data
    - tidy and normalize data
    - subset and group
    - test hypotheses
    - create intuitive plots that include test statistics
  - Save script as readable report