**R COMPANION OUTLINE**

* *Italics* are section heads from the book
* Non-italic font is from me
* pink indicates that this section will likely have no correspondence to the companion, but it is included for clarity of the outline

**Chapter 0**

1. What is the purpose of this companion/book?
   1. accompanies OI Biostat
      1. explain the difference between the two
      2. conceptual vs. analytical
   2. an introduction to computing using R, particularly for use in the life sciences
   3. to build an understanding of how data is stored, processed, transformed, and analyzed
2. What is R?
   1. how to install it
   2. recommend that you use RStudio instead of just R
   3. what does RStudio look like?
      1. explain the usage of the various parts
   4. how to create a R script and why to use that rather than the console
   5. how to import a dataset into R
3. how does R think of data? What are the basic data types?
   1. vectors
   2. matrices
   3. dataframes
   4. lists
   5. small examples of each
      1. how to create them
      2. how to name them
      3. how to access elements within them (discussion of bracket notation and rows v columns etc)
      4. how to manipulate, combine, transform them
   6. explain relationship between vectors and matrices, datasets, dataframes and the differences therein
   7. Creating a new variable from others (both generally and in the dataset itself)
   8. data types
      1. numerical, factor, etc
      2. how to convert between them?
      3. how to check what the data type is
      4. add some notes about how sometimes when you upload a dataset to R the data is misclassified so you should check and convert if need be
4. Basics of using functions in R
   1. explain how all functions will be introduced in this text
   2. idea of inputs inside parentheses
   3. help on function commands
5. Basic programming background
   1. a for loop
   2. understanding boolean
   3. understanding logic operators, such as >, <, ==, !=
   4. if statements
   5. again going back to pointing out bracket notation here
   6. the concept of subsetting
6. The R package we have created
   1. a resource of datasets to provide examples
      1. say that they can follow along through the examples provided in the text using this package
      2. a good place to start to learn how R packages work, etc
   2. how to download it from CRAN (or our yet created website?)
   3. how to use it once you have it
   4. help(package name) — explain all the elements of a dataset in R
      1. note that not all R datasets will be from a package so they might not have this much information
7. Exercises
   1. Two data problems
      1. Opening a dataset from the package
      2. Uploading your own dataset (we can link to one available online somewhere)
   2. Accessing certain parts of a dataset
      1. bracket notation, subsetting, etc
   3. Creating a new variable/manipulation of current variables into a new one
   4. Ex. find how many people in the study have xx trait
   5. small programming examples
      1. create a for loop and print out the numbers 1 though 10
      2. create a for loop and print out only even numbers
      3. print out every other element in a list of names

**Chapter 1**

1. *Case study: preventing peanut allergies* 
   1. how to access a dataset in the package reminder
   2. show how to obtain Table 1.1
      1. do this multiple ways to illustrate various methods
   3. introduction to table command
      1. also show Table 2.1
      2. include with and without add margins
2. *Data Basics* 
   1. *Observations, variables, and data matrices*
   2. *Types of variables*
   3. *Relationships between variables*
3. *Data Collection Principles*
   1. *Populations and Samples*
   2. *Anecdotal Evidence*
   3. *Sampling from a population*
   4. *Sampling Methods* 
      1. Introduce the use of sampling in R here
      2. We would be able to do both simple random sampling as well as stratified sampling (would require for loop and if statements in comparison to just sample command)
   5. *Introducing Experiments and*  observational studies
   6. *Experiments*
   7. *Observational Studies*
4. *Numerical Data*
   1. *Measures of center: mean and median* 
      1. introduce summary command above this section, reference 5 number summary
      2. show how to calculate mean both by hand and using R command
      3. includes mention of several other functions: length, sum, round
      4. median command in R
   2. *Measures of Spread: standard deviation and interquartile range* 
      1. again show by hand and with R command
      2. var, sd, iqr commands
      3. lay out clearly what each one takes as input and option to include na.rm = TRUE
   3. *Robust Estimates* 
      1. an example of using boolean with real datasets
      2. I think this may be a better spot in the companion to explain taking summary statistics of part of a dataset, i.e. subsetting or logical commands, rather than robustness or not since robustness is just a definitional issue not an analytical one
   4. *Visualizing distributions of data: histograms and box plots* 
      1. histogram command
         1. show comparison of changing certain features
      2. box plot
   5. *Transforming Data* 
      1. log before and after side by side plots
5. *Categorical Data*
6. *Relationships between two variables* 
   1. *Two numerical variables* 
      1. *scatterplots*
         1. plot command
         2. one or two examples of a super basic one versus adding a few features (maybe side by side)
      2. *correlation*
         1. is it worthwhile to calculate the correlation by hand??
         2. the cor command
   2. *Two categorical variables* 
      1. frequency table
      2. *contingency tables* 
         1. both a contingency table and a proportion table (in both row and column directions)
      3. bar plots
      4. *segmented bar plots*
         1. how to do this
         2. a standardized segmented bar plot
      5. *two by two tables and relative risk*
   3. *a numerical variable and a categorical variable* 
      1. side by side box plots
      2. overlapping histograms
7. *Exploratory Data Analysis* 
   1. *case study: discrimination in developmental disability support*
   2. *case study: Molecular cancer classification*
   3. *case study: cold-response genes in the plant Arabidopsis arenosa*
8. *Notes*
9. *Exercises* 
   1. perhaps modeled off of the ones in the book, for example, pulling the data, creating the tables used there and then solving for the answers
   2. looking at a dataset and determining how a data category is classified in the dataset
   3. taking a sample of a dataset
   4. looking at the dataset, plots, summaries, etc

**General thoughts**

* standardize the introduction of new functions
  + i.e. here is the function
  + here are the possible inputs and what form they should be in
  + here are the outputs that come out
    - and different types of outputs you can get
  + then show an example of it (possibly more than one to demonstrate how different features change the outcome)