Technology Plan

Objectives: Create a statement for students that lets them know what technology they will need to have and what technology skills they will need to be successful Make technological choices that will serve students well as their careers progress

Task: Write a technology plan for the distance/hybrid course you plan to teach. This should explain to students how technology will be used to support instruction and learning in your course. What technologies will you use? What support do you and your students need to be successful? Is there a YouTube tutorial that will help your students get started?

Technology is a central component of data science, computer science and some applied mathematics courses whether they are distance/hybrid or face-to-face. This subject has been a particular challenge to me for many years. There are two reasons for this: (1) use of computational tools instead of manual calculations can inhibit development of students' mathematical skill and understanding and (2) making software available to students requires computing resources and technology support (e.g., someone to install and maintain the software assuming not all students can do this on their own). As a graduate student at the University of Illinois, I taught in the Calculus & Mathematica program. This was a discovery-based course in which students used *Mathematica*, a powerful, commercial software package, to work through calculus laboratory activities in class. See www.wolfram.com/mathematica. The associated lessons were written by research mathematicians at Illinois and Ohio State. They included deep insights and innovative graphics to illustrate concepts. I found that students in this course, however, were extremely weak in their ability to do calculations by hand even for trivial calculus problems. As a result, I began including a significant hand-written component to my courses. Over the years I have seen a significant decline in students' ability to compute with fractions and solve simple equations without resorting to a calculator. (I'm currently fighting that battle at home with my 7th graders!)

After leaving Illinois for my first teaching position, *Mathematica* was no longer available to me (its license fee is high). Similarly, when I worked in industry, it was effectively not available to me (we had a small number of licenses at a 100+ employee company). My experience with other commercial software products is similar. In universities and business, I have been required at different times to use Matlab, Minitab, Maple, SAS, Microsoft Office and the Microsoft Windows operating system. In every instance, that required tool was not available to me (unless I wanted to pay for it personally) in some subsequent position.

I am a strong advocate for the use of open-source software. Not only is it consistently available to me in diverse work environments, I have found it to be a more reliable work partner. In the classroom, I try to provide students with opportunities to use computational tools that they will have access to throughout their careers rather than tools that a commercial vendor makes available to academics so that users will pay license fees in the future (or collect data on us now while we are using their products). The tools that I use include the Linux operating system rather than Windows or Mac, LaTeX rather than commercial word processors, and data science and mathematics applications such as the R environment for statistical computing, programming languages such as Python and Java, the PostgreSQL database, the QGIS geospatial mapping application and others. These are also the tools that I turn to when working on research and consulting projects.

Institutional and corporate computing services tend not to provide technical support for open-source tools. At one company where I worked, the attitude of IT seemed to be "Who will we sue if we get hacked?" As a result, I have spent many hours over the years learning to provide my own IT support with open-source tools. It has been challenging but interesting and of enduring value in my career. At Shepherd University, I managed a computer lab where I could ensure students had access to course software (at least when they were in the lab). At Shenandoah, it is a great benefit that students all have their own laptops. Understanding the unique issues that can arise in installing software (e.g., R and PostgreSQL which are not supported by IT) on particular students computers can be a challenge for me, however.

In Math 207 I use the R environment for statistical computing (https://www.r-project.org) and the RStudio graphical interface (https://www.rstudio.com). I have been using these in similar courses since 2011. Last year at Shenandoah I experimented with project assignments using R and a the National Health and Examination Survey (NHANES) data set (https://www.cdc.gov/nchs/nhanes). An example is included at the end of this assignment. Some students found these quite challenging and a few seemed frustrated. However, several other students took quickly to these projects and wrote solutions above and beyond what was required. For an online course, it will be especially important for me to develop examples and 'smaller steps' that allow more students to make rapid initial progress.

- In the first week of class, I have students install R and R Studio. My instruction sheet is included on the next page.
- Once students have the software installed, I have them complete a worksheet that illustrates basic use of R as a calculator, graphing tool and platform for doing statistical computations and running simulations (e.g., flip 10 coins 1000 times and make a histogram of the results).
- Within the first month, I assign a project in which students read the NHANES data set into R then compute descriptive statistics and plot histograms of specified variables and variables of interest to them (the data set contains thousands of variables recorded for over 30,000 subjects). For an on-campus course, I have students generate pdf reports as outputs to their labs. This requires installation of another software application that R uses. For an online course, I will have students turn in their R files instead since my ability to help them install software will be more limited.
- As the semester progresses, I will assign R worksheets once or twice per week that show students the R commands related to our class concepts and allow them to experiment with examples.
- I also plan to continue assigning R projects with the NHANES data set. These start with examples that I include in the R lab file. Students are asked to write comments responding to questions in the file (e.g., The correlation between age and systolic blood pressure measurements among subjects in the study was 0.15. Comment on the sign and magnitude of this value.). As the semester progresses, I ask more challenging questions (e.g., Select two quantitative variables of interest to you. Compute the regression line, plot the data and the regression line and and comment on the results.)
- For these R projects, I will have students turn in their R files (rather than a pdf) on Canvas.
- I have prepared the full NHANESIII data set as a zip file that students can download from Canvas.

Math 207
Introduction to Statistics

Installing R and RStudio

The R Environment for Statistical Computing is a free software environment for statistics and graphics. RStudio is a graphical user interface to R. Both are used extensively in academics, industry and government. They are freely-available to download for Mac, Linux and Windows. We will use both in this course. Please follow the steps below to install and test them on your computer.

Install R:

- Open a web browser to https://www.r-project.org.
- Click on the *download R* link near the top of the page.
- Scroll down the CRAN Mirrors page and select any U.S. site.
- Click on the *Download for (Mac) OS X* link (or the link that is appropriate to your computer's operating system).
- Once the download is complete, install the software. On Mac, this involves double-clicking on the downloaded image file.

Install R Studio:

- Open a web browser to https://www.rstudio.com.
- Click on the *Download RStudio* link near the top of the page.
- Click on the RStudio Desktop, Open Source License, FREE Download link.
- Click download for your computer's operating system (e.g., Mac OSX).
- Once the download is complete, install the software. On Mac, this involves double-clicking on the downloaded image file.

Here is a video that demonstrates installing R and R Studio on Mac OS X:

https://www.youtube.com/watch?v=GLLZhc_5enQ

RStudio Cloud: As an alternative to installing and running R and RStudio on your machine, you can use the RStudio Cloud interface. Just open

https://rstudio.cloud

in a web browser, sign up for an account and start using RStudio!

Relationship to the International Society for Technology Education Standards

1. **Empowered Learner**. Students leverage technology to take an active role in choosing, achieving and demonstrating competency in their learning goals, informed by the learning sciences.
2. **Digital Citizen**. Students recognize the rights, responsibilities and opportunities of living, learning and working in an interconnected digital world, and they act and model in ways that are safe, legal and ethical.
3. **Knowledge Constructor**. Students critically curate a variety of resources using digital tools to construct knowledge, produce creative artifacts and make meaningful learning experiences for themselves and others.
4. **Innovative Designer**. Students use a variety of technologies within a design process to identify and solve problems by creating new, useful or imaginative solutions.
5. **Computational Thinker**. Students develop and employ strategies for understanding and solving problems in ways that leverage the power of technological methods to develop and test solutions.
6. **Creative Communicator**. Students communicate clearly and express themselves creatively for a variety of purposes using the platforms, tools, styles, formats and digital media appropriate to their goals.
7. **Global Collaborator**. Students use digital tools to broaden their perspectives and enrich their learning by collaborating with others and working effectively in teams locally and global

The use of R and, in particular, projects involving real data sets, do seem to align with certain standars of the International Society for Technology Education.

- Having students select variables of interest in the NHANES data set, perform analysis and discuss their findings aligns with standard 1.
- Use of a large, real-world data set aligns with standard 5.
- Having students interpret numerical and graphical results in projects and smaller assignments aligns with standard 6.

Report_2

Your Name

16 October 2019

Reading the Data Files

The NHANES data set consists of four data files: adult.csv, youth.csv, lab.csv and exam.csv. The adult file contains information about subjects who were over 17 years old. The youth file contains information about younger subjects. The lab and exam files contain additional data about both adults and youth.

Reading the adult.csv file

The following command reads the adult.csv file into R.

```
dataDir = "/home/ralphw/su2/math207/nhanes/"
adult = read.csv(paste(dataDir,"adult.csv", sep=''), header=TRUE)
```

The size of this data set is

```
dim(adult)
```

```
## [1] 20050  1238
```

This means that the file has 20,050 rows and 1238 columns. So, there are 20,050 adults in the data set and 1238 pieces of information about each in this file. The lab.csv and exam.csv files include additional information about these subjects.

Reading the youth.csv file

The following command reads the youth.csv file into R.

```
youth = read.csv(paste(dataDir, "youth.csv", sep=''), header=TRUE)
```

The size of this data set is

```
dim(youth)
```

```
## [1] 13944   687
```

This means that the file has 13,944 rows and 687 columns.

Reading the exam.csv file

The following command reads the exam.csv file into R.

```
exam = read.csv(paste(dataDir, "exam.csv", sep=''), header=TRUE)
```

The exam of this data set is

```
dim(exam)
```

```
## [1] 31311  2368
```

There are 31,311 rows and 2368 columns in the exam file.

Reading the lab.csv file

The following command reads the lab.csv file into R.

```
lab = read.csv(paste(dataDir, "lab.csv", sep=''), header=TRUE)
```

The size of this data set is

```
dim(lab)
```

```
## [1] 29314   356
```

There are 29,314 rows and 356 columns in the lab file.

Regression Examples

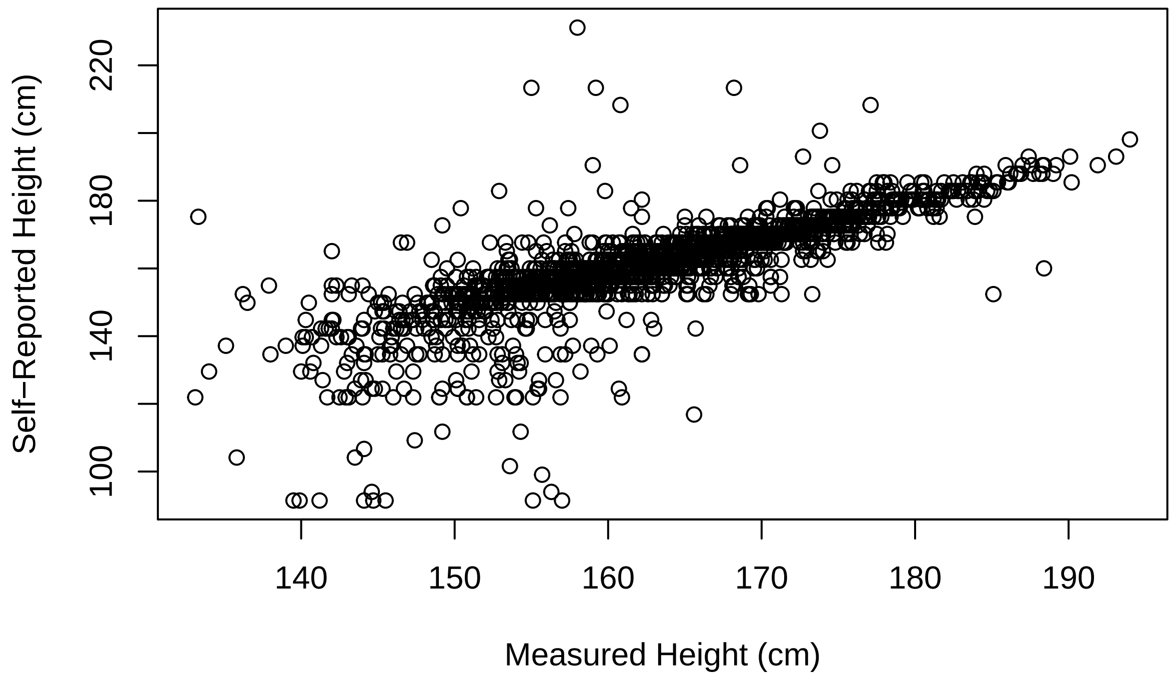
Measured Height vs Self-Reported Height

Extract the subject ID, measured height and self-reported height. Then remove the NA (blank) values and remove the missing values (error codes 88888 and 888). Finally, convert self-reported height from inches to cm. See page 199 of the exam-acc.pdf code book.

```
x1 <- data.frame(SEQN=exam$SEQN, BMPHT=exam$BMPHT, BMPSRHIS=exam$BMPSRHIS)
x2 <- na.omit(x1)
x3 <- x2[(x2$BMPH != 888888) & (x2$BMPSRHIS != 888), ]
x3$BMPSRHIS <- 2.54 * x3$BMPSRHIS
```

Plot the measured and reported heights.

```
plot(x3$BMPHT, x3$BMPSRHIS, xlab="Measured Height (cm)",
     ylab="Self-Reported Height (cm)")
```



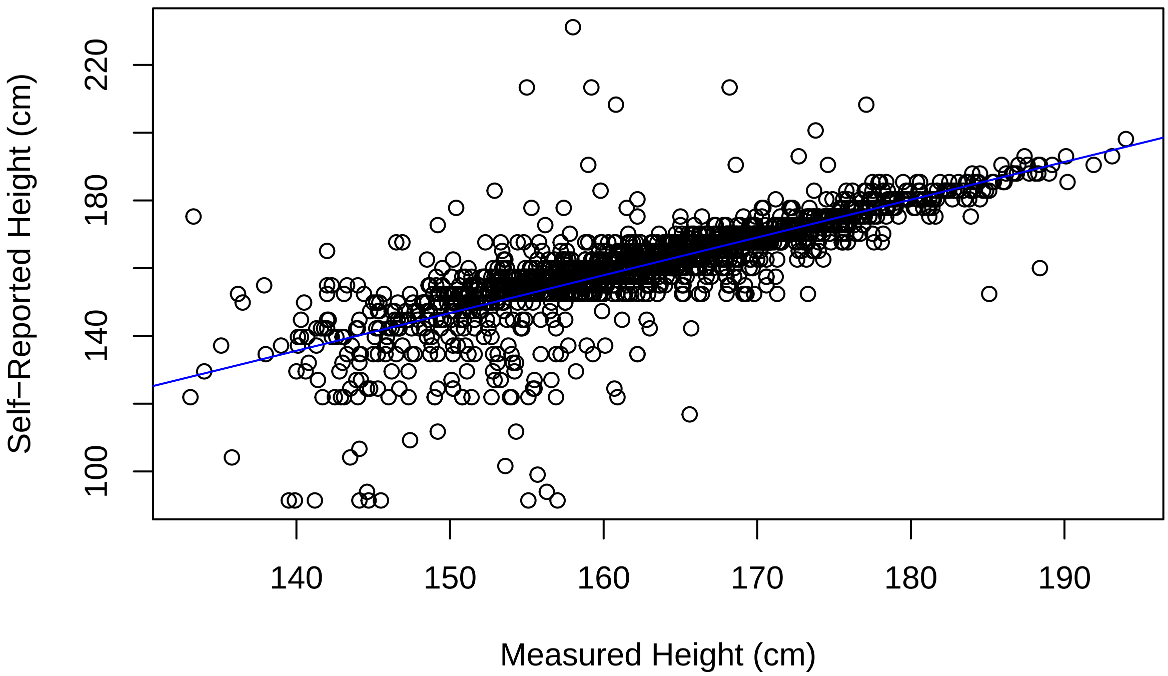
The correlation appears to be strong and positive.

```
cor(x3$BMPHT, x3$BMPSRHIS)
```

```
## [1] 0.7819654
```

Now compute the regression line and add it to the plot.

```
model = lm(x3$BMPSRHIS ~ x3$BMPHT)
plot(x3$BMPHT, x3$BMPSRHIS, xlab="Measured Height (cm)",
     ylab="Self-Reported Height (cm)")
abline(model, col='blue')
```




```r
summary(model)
```

```
##
## Call:
## lm(formula = x3$BMPSRHIS ~ x3$BMPHT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.133  -1.515   0.740   2.896  75.452
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.4829     3.2745  -6.255 4.86e-10 ***
## x3$BMPHT      1.1150     0.0201  55.469  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.477 on 1955 degrees of freedom
## Multiple R-squared:  0.6115, Adjusted R-squared:  0.6113
## F-statistic:  3077 on 1 and 1955 DF,  p-value: < 2.2e-16
```

From the output we determine that the regression line is

$$\text{self-reported height} = 1.11 \times \text{measured height} - 20.48$$

This equation suggests that taller people in the study tended to overestimate their heights.

Measured Height vs Recumbent Length

Do a similar regression calculation to compare BMPHT and BMPRECUM.

Measured Height vs Weight

Do a similar regression calculation to compare BMPHT and BMPWT.

Measured Height vs Head Circumference

Do a similar regression calculation to compare BMPHT and BMPHEAD.

Age vs Body Mass Index

Do a similar regression calculation to compare HSAGEIR and BMPBMI.

White Blood Cell Count vs Red Blood Cell Count

Do a similar regression calculation to compare WCPSI and RCPSI in the lab data set. See pages 70 and 71 of the lab-acc.pdf code book.