Why, When,What & How R? -& the training of teachers

Edith Seier PhD Statistics

Professor Emerita ETSU

seier@etsu.edu

http://faculty.etsu.edu/seier

Influence of what we do in what we teach

- When we practice professionally has an influence in what we teach.
- This is particularly true for statisticians
- And even more when you teach programming, we tend to include what has been useful for us.
- Thus, teaching material prepared by two people to teach the same language might make emphasis on different commands and use different examples.
- We need to involve high school teachers if we prepare teaching material for high school students.

My experience with computers

- Started with Fortran IV and punched cards ☺ before personal computers & Windows existed. In the mainframe with terminals had to use more tan one operating system. My main motivation: be able to do research in Statistics
- Consulting for a 1-12 School ~ 1990 about the use of computers that were just being introduced in the school, we started with Commodore computers and used LOGO as the language of choice for students (LOGO was developed by Papert, Solomon & Feuzerg as an educational programming language (S. Papert was an expert in Education & Cognitive Science from MIT)
- For short periods used Timeslab for time series & Lisp-Stat
- In the past I have used mainly GAUSS, but also MATLAB for my research, and now I use mainly R! (for applications: R, SAS & Minitab)
- Taught several Stats courses at ETSU using statistical software. Designed & taught the statistical computing course in the Math&Stats department 1/3 SAS & 2/3 R
- Now trying to learn more R with online courses & some Phyton

Why programming and not only canned programs?

- It makes our students think, what is the goal, how to disentangle a problem in little pieces, pay attention to details to fix errors.
- They have to really understand something in order to tell the computer how to do it
- Programming is a skill required in many jobs/professions now, students deserve at least to know what it is about.
- It might open their eyes to the existence of career paths they have not been aware before (statistics, business analytics, data science,……)
- Maybe the most important: You not only learn something you learn a tool to learn other things

Knowing how to program can make a difference (Remember the movie 'the social network'? What was the difference between Mark Zuckerberger and the Winklevoss brothers? ☺)

Why R?

- It is free. The thought that anybody who has access to a computer in any part of the world can use R legally is wonderful!
- There is a nice community who writes new packages and tutorials and puts them for everybody to use, a real lesson on sharing
- Does everything canned programs can do & more
- Has nice graphic capabilities
- Anything that is done with graphic calculators could be done with R
- We can teach the basics and there are many opportunities on line to learn more for those who wish to do it (Coursera, Data Camp,...)

What to learn in R?

Operating systems and languages might pass but the experience of thinking to be able to write code, solve problems and fix errors remains with you

What we learn about a programming language depends of what we want to do. Usually we learn what we need. That is easier to decide when you are already in a job or profession but more difficult to tailor when we don't know yet what the students are going to do in life as happens in high school. Who knows what language will be fashionable in the future

My opinion:

- some basic data analysis including graphs (because we all need that) &
- some tools useful to make us think & explore knowledge (loops, conditional statements & writing functions).

Programming is a good exploration tool

Two examples to show how programming how to explore problems before students have the theoretical knowledge that solve them

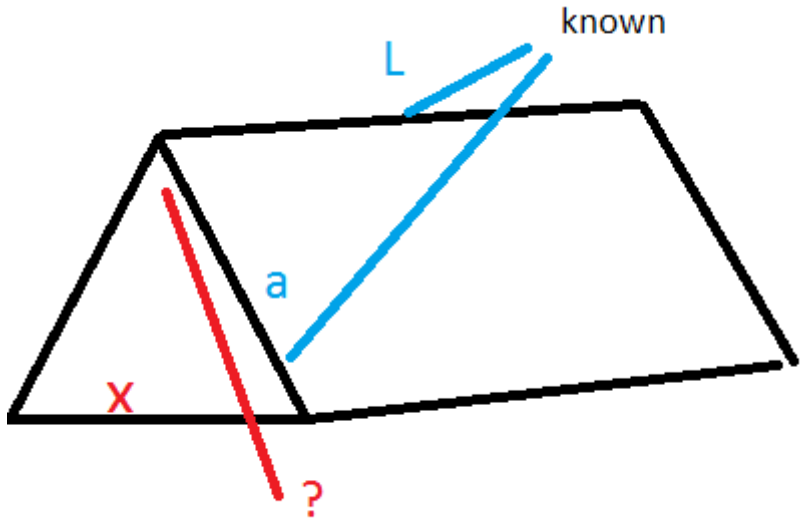
(Math) Not all students get to take Calculus in High School therefore they might not learn about maxima and minima problems.

(Stats) One of the topics that is more difficult to grasp for students in the Intro Stats course is sampling distributions.

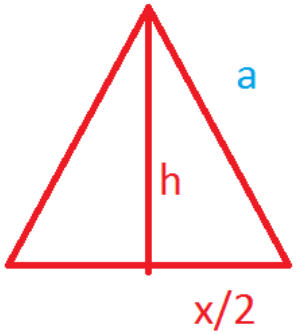
Lets see what R can do for us in that regard.

The classical tent problem

The question is : What is the angle that maximizes the volume

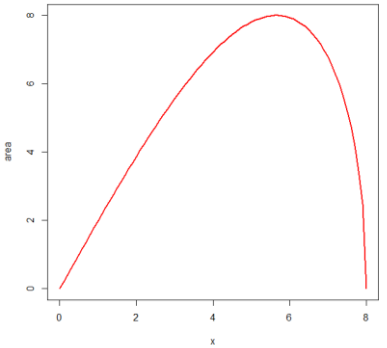


Simplifying the problem:
Since the length L is known, maximizing the volume is equivalent to maximizing the area of the triangle.



Area= (x * h)/2
but using Pythagoras a^2=h^2 +(x/2)^2
thus $h = \sqrt{a^2 - x^2/4}$

So we can explore how the area of the triangle behaves as we make x vary (and h varies as well) for a given known a.
First lets assume that a is 4 yards
Let x vary between 0 and 8 (the extreme possible cases)

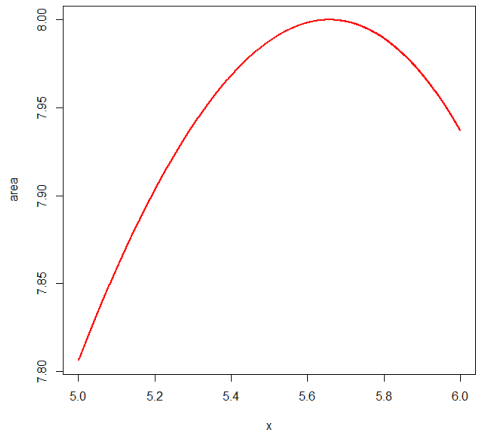



```
x<-seq(0,8,by=0.1)
h<-sqrt(4^2-(x^2)/4)
area<-(x*h)/2
table<-cbind(x,x/2,h,area)
plot(x,area,'l',lwd=2,col='red')
table
max(area)
```

```
[55,] 5.4 2.70 2.9512709 7.9684315
[56,] 5.5 2.75 2.9047375 7.9880282
[57,] 5.6 2.80 2.8565714 7.9983998
[58,] 5.7 2.85 2.8066884 7.9990621
[59,] 5.8 2.90 2.7549955 7.9894868
[60,] 5.9 2.95 2.7013885 7.9690962
```

Lets further explore that interesting region

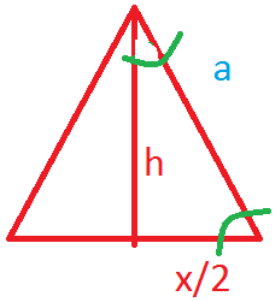
	x	x/2	h	area
[55,]	5.4	2.70	2.9512709	7.9684315
[56,]	5.5	2.75	2.9047375	7.9880282
[57,]	5.6	2.80	2.8565714	7.9983998
[58,]	5.7	2.85	2.8066884	7.9990621
[59,]	5.8	2.90	2.7549955	7.9894868
[60,]	5.9	2.95	2.7013885	7.9690962



We notice that h is very similar to x/2

Lets work now with a finer grid in the region5<x<6

```
x<-seq(5,6,by=0.001)
h<-sqrt(4^2-(x^2)/4)
area<-(x*h)/2
table<-cbind(x,x/2,h,area)
plot(x,area,'l',lwd=2,col='red')
table
max(area)
```



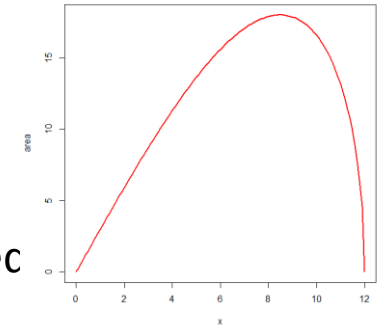
The maximum area is 8 and it happens when x/2 and h are equal

[658,] 5.657 2.8285 2.828354 8.000000

So if h and x/2 are equal that means that the two angles have to be the same, since they have to add 90 degrees, that means each is 45 and the top angle is 45x2=90

OK, that happens when a=4, but,..does it happen in general? The students can explore how happens for other values of a and at the same time think on how to generalize something and learn how to write a function in R by recycling the commands we had used before.

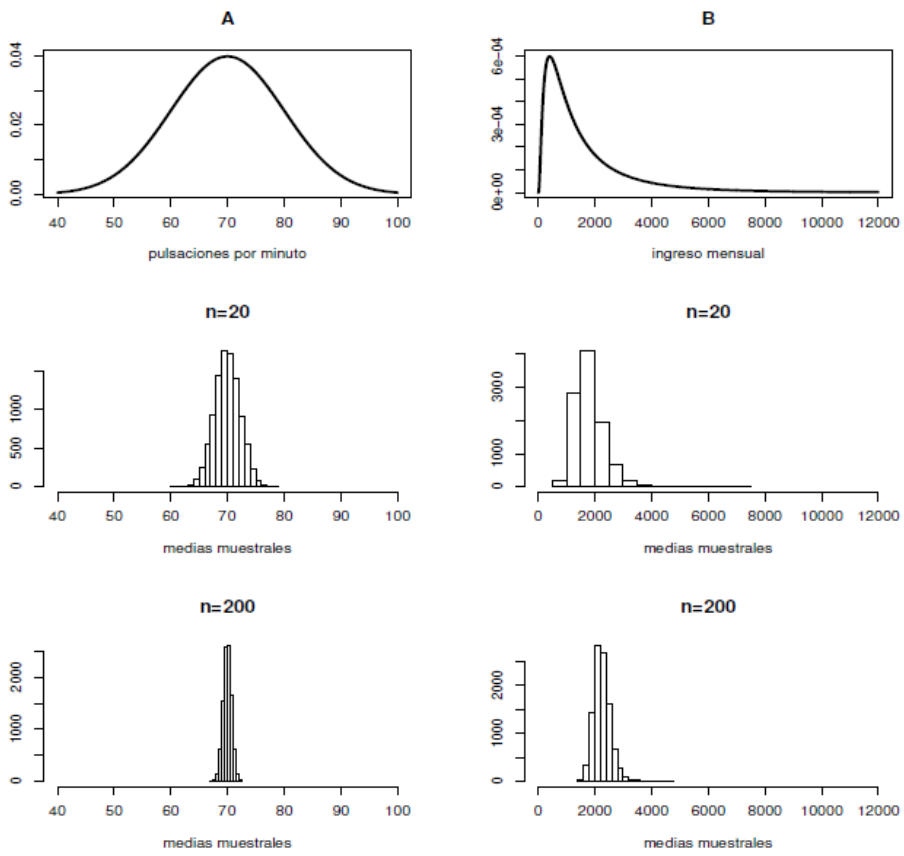
```
tent<-function(a)
{
x<-seq(0,2*a,by=0.1)
h<-sqrt(a^2-(x^2)/4)
area<-(x*h)/2
table<-cbind(x,x/2,h,area)
plot(x,area,'l',lwd=2,col='rec
table
}
```



To do the calculations for a=6, just write

```
tent(6)
```

An example in Stats: Sampling distributions



We learn about sampling distributions but also we learn about loops

```
par(mfcol=c(3,1))
pulso<-40:100
dpulso<-dnorm(pulso,70,10)
plot(pulso,dpulso,'l',lwd=2,xlab='heartbeats per minute')

xbar20<-numeric(10000)
for(i in 1:10000) {
x<-rnorm(20,70,10)
xbar20[i]<-mean(x)
}
hist(xbar20,main='n=20',xlab='sample means', xlim=c(40,100))
xbar200<-numeric(10000)
for(k in 1:10000) {
y<-rnorm(200,70,10)
xbar200[k]<-mean(y)
}
hist(xbar200,main='n=200',xlab='simple means',xlim=c(40,100))
```

How?

- Hands on
- Exercises
- Preferably motivated by a problem to maximize the ‘thinking’
- Class discussion
- Work with your neighbor
- Multiple Choice Quizzes
- Personal projects

When?

- At College level the natural place is a Statistics course but can be introduced in other courses as well
- In High School, Math, Stats or Physics courses would be the most convenient if there is not a programming course. For anything that a calculator is used nowadays, R could be used if computers are available.
- Work has to be done to select appropriate topics according to the course in which R is going to be ‘inserted’ as well as teaching material. The participation of high school teachers in suggesting topics (of math or other courses) that could be used as ‘seeds’ to develop teaching material in R is fundamental.
- Information should be given about the opportunities that knowing how to program opens and the resources that exist out there to continue learning.

Training of teachers

- It takes some time between we learn something and we are able to teach it and create teaching material
- Thus, when we teach R to instructors without previous experience with R we should not only teach R itself but provide some guidance about the way to teach it and some teaching material.
- It is not the same to teach a course to teachers than a course for students. The objective of each topic and the thinking process associated with exercises and projects need to be discussed.

The ‘Learning and Teaching R’ Workshop at Universidad de San Marcos, Lima, Peru, March 2017

- It took place in a department where several of the instructors are enthusiastic users of R, have learned it on their own and use it in the upper division courses. In those courses the emphasis is put in how to apply the statistical methods studied in the course not in the programming itself
- The students majoring in Statistics learn R in the statistical computing course (junior level course) and they love R (they won second place in a local DataFest competition) .
- But not all of the older instructors had learned or use R and many use SPSS etc. The workshop was designed for them. However, several of the instructors knowledgeable in R also attended.
- We started from 0, in an environment where they could feel comfortable to ask questions etc. Spontaneously some of the participants with experience in R helped those who were tending to stay behind.
- Mainly based on the course STAT2090 Statistical Computing at ETSU but discussing with the participants why we include those topics as well as teaching material such as quizzes, projects etc.
- 3 mornings (MWF 10-1) in a computer lab. Printed class notes in Spanish and online scripts with the commands were provided to save time.

The content of the workshop

- Starting to work with R
- More on graphics
- Transforming variables
- Loops
- Writing functions
- Probability and other Math topics
- How to continue learning

Final thoughts

- To learn a programming language is not too different from learning a foreign language, after a basic vocabulary we learn from seeing programs and practicing.
- Even when the language is the same the way we teach it and the examples used has to be thought thinking on the specific audience
- To learn R might not be the most important goal for high school teachers and students, even if it is a very good one
- We want to give high school students a flavor of what programming is and in that way to open a window for them into an activity that is now everywhere in the real world
- We also want to provide them with an exploration tool that can be used to learn other things (as in the examples we saw)
- It is also an opportunity to provide information & motivation to students (maybe inviting people who work in the industry and use programming to give talks) to consider careers that involve programming
- The participation of high school teachers in the development of teaching material adequate for high school students is fundamental