SASnatch: Using SAS Naturally

(or knitrally)
(or snatchurally)

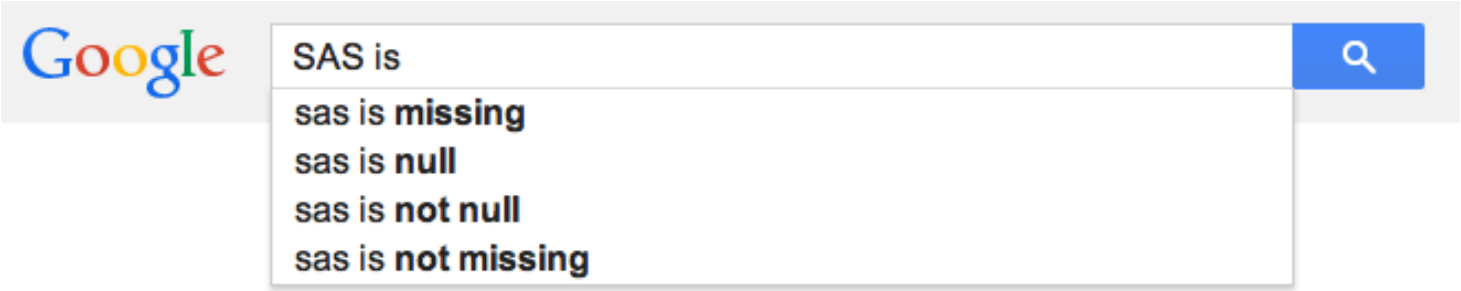
Ian Mouzon

STAT 585 Presentation
Spring 2014

April 30, 2014

What is SAS

And why should you care



Multiple Choice: SAS is one of the following

(a) SAS is a software suite developed by SAS institute for advanced analytics

(b) A frustrating and cumbersome tool that has been losing its place in academia and increasingly in industry and which will not be missed when it disappears forever

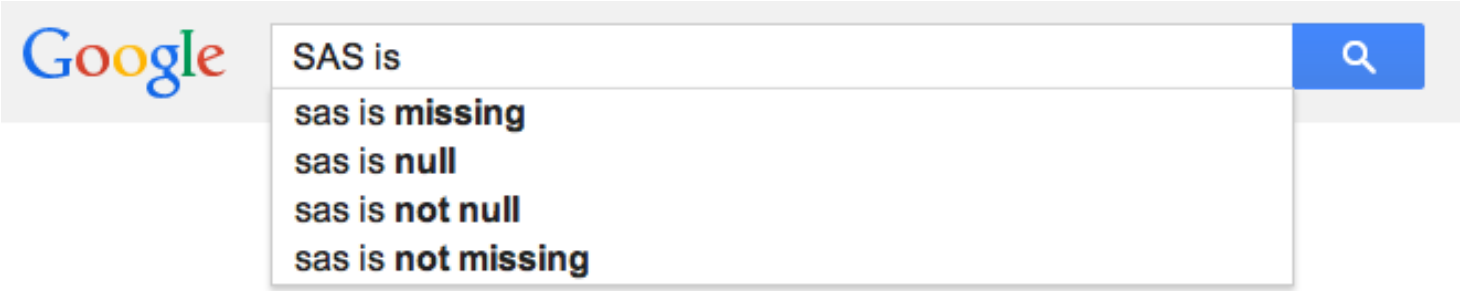
(c) A programming language that you will have to use in one or two classes but then all your professors after that will use R

(d) All of the above

The answer is (a). The other choices are all incorrect.

What is SAS

And why should you care



Multiple Choice: SAS is one of the following

(a) SAS is a software suite developed by SAS institute for advanced analytics

(b) A frustrating and cumbersome tool that has been losing its place in academia and increasingly in industry and which will not be missed when it disappears forever

(c) A programming language that you will have to use in one or two classes but then all your professors after that will use R

(d) All of the above

The answer is (a). The other choices are all incorrect.

SAS: It's actually really good

It's better than you remember and does more than you knew

- SAS code that works now will still work years from now and SAS code that ran years ago will still run today
- SAS is very well documented (something that matters more as you move away from STAT 500 material),
- **Three of the big topics from this class**
  - SAS has interactive graphics,
  - built in tools for *static or dynamic* web publishing,
  - built in database management methods
- SAS runs of linux

Example: Principle components

```
DATA TURTLES;
   INFILE "./turtles.dat";
   INPUT LENGTH WIDTH HEIGHT SEX;

PROC PRINCOMP COV OUT=PCA1;
   VAR LENGTH WIDTH HEIGHT;

PROC PLOT DATA=PCA1;
   PLOT PRIN2*PRIN1;
   PLOT PRIN3*PRIN1;
   PLOT PRIN3*PRIN2;
RUN;
```

A few problems you may have just noticed

Or noticed over and over again in STAT 500

- There is too much output.
  - SAS seems to err on the side of caution and gives us more than we usually need.
- The plots are underwhelming
  - SAS can create very high quality plots but these don't seem to be
  - We can always export the data sets so that we can make our plots in R, but this would be an enormous amount of trouble
- The output does not allow us to make changes.
  - We can dig into the html file produced in the second example but this is a far cry from the ease of using knitr

Most of my complaints about SAS have to do with the way it fails to work like knitr.

A few problems you may have just noticed

Or noticed over and over again in STAT 500

- There is too much output.
  - SAS seems to err on the side of caution and gives us more than we usually need.
- The plots are underwhelming
  - SAS can create very high quality plots but these don't seem to be
  - We can always export the data sets so that we can make our plots in R, but this would be an enormous amount of trouble
- The output does not allow us to make changes.
  - We can dig into the html file produced in the second example but this is a far cry from the ease of using knitr

Most of my complaints about SAS have to do with the way it fails to work like knitr.

SASnatch: Using SAS in a more natural way

SASnatch: The Power to Snatch

SASnatch is a package that aims to do the following:

- Transfer data between SAS and R with *minimal* effort,
- Eliminate clicking around SAS, and
- Manage SAS's output in a convenient way

Which will hopefully

- Allow the programmer to use which tool he or she prefers at that moment, be it is SAS or R
- Make SAS more accessible to R users

It does this through knitr (generally). I will show you how this works.

Next steps punchlist

Originally "easy, medium, and hard"

R:

- Incorporate SAS's log file into the inline call (i.e., `cat(log.file)`)
- Improve "chunk" behavior (the code should print, but must use `eval = FALSE` ...)

LaTeX:

- Syntax highlighting the SAS code in the output document (LaTeXpackages can do this, but using them with knitr is tricky)
- Using the SAS's higher level LaTeXoutput (which would require accessing and using SAS's "style sheets")

SAS:

- Syntax highlighting while programming (VIM can kind of do this)
- Improve data management using databases (SAS and R can both read and write from SQLite - but it's not on the terminals)

Questions?

Thank you for listening!

SASnatch on github at: `github.com/imouzon/SASnatch`

Installation:

```
>library(devtools)
>install_github('SASnatch','imouzon',
                arg='-l U://Documents/R/win-library/3.0')
```