

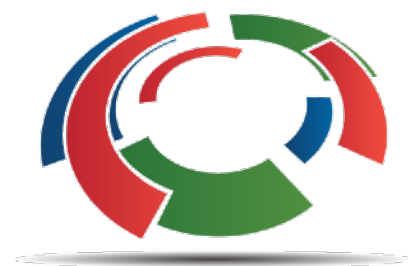
Slides here...

<http://ajstewartlang.github.io/CarpentryConnect2019/slides.pdf>

Basic Statistical Models Using R

Andrew Stewart and Peter Smyth

 @ajstewart_lang



Software
Sustainability
Institute

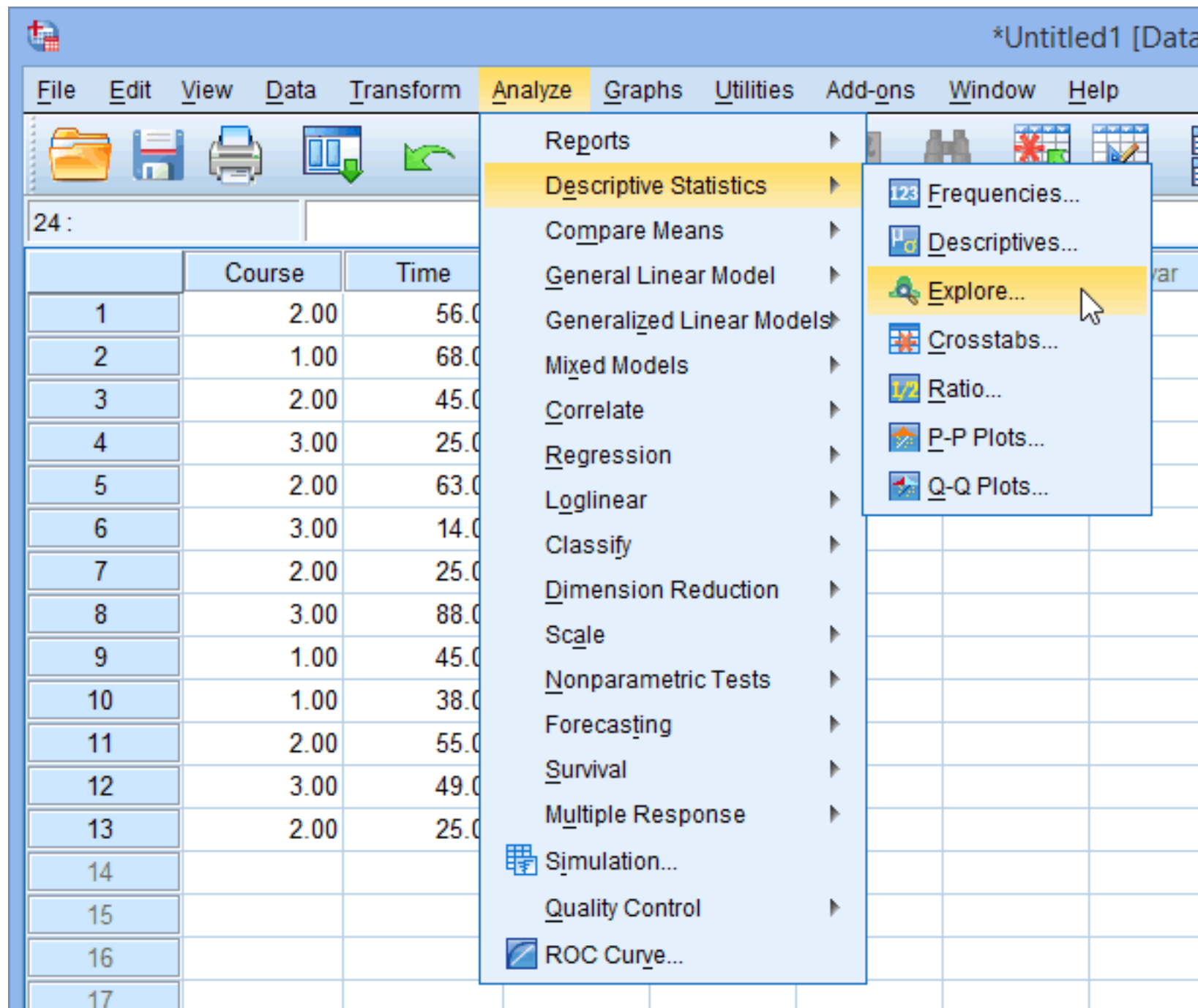


© DAILY MAIL

The Challenges

- Teaching Statistics and R to M-level students in Psychology - 2 units on MRes and new MSci unit.
- Mixed background and interest in Statistics.
- Mixed background and interest in coding (with many never having coded before).
- Most people's UG backgrounds are similar...

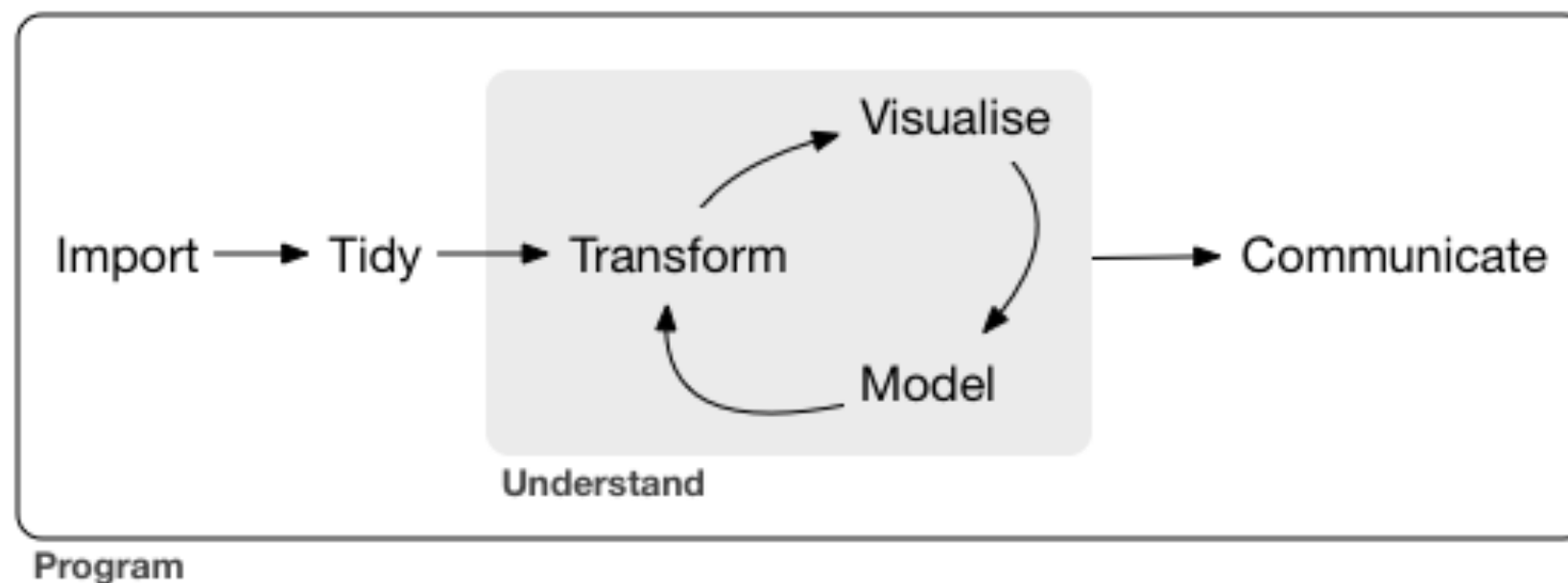
Pointy and Clicky



The problem is that you can't really do reproducible analysis in a GUI...

And it's not so much as statistical knowledge, but knowledge about what to click (and where)...

- Reproducibility is key in science - and one of the easiest ways to engage in reproducible research is to use an open source statistical language such as R.
- The tidyverse workflow allows for data importing, wrangling, visualisation, and modelling all in the same reproducible workflow.



Hadley Wickham and Garrett Grolemund

What we'll cover in this session...

- Data simulation
- Data visualisation
- t-tests
- General linear model (continuous predictors)
- General linear model (factorial)

What packages we'll use in this session...

```
install.packages("tidyverse")  
install.packages("broom")  
install.packages("afex")  
install.packages("emmeans")
```

```
library(tidyverse)  
library(broom)  
library(afex)  
library(emmeans)
```


Testing for
differences
between groups
using Student's t-
distribution.



Let's create some data...

- Imagine we have two groups of participants - 50 in each group. We are interested in the effect of caffeine on reaction time. Sample A we give an espresso, Sample B a de-cafeinated coffee. We measure the reaction time of each individual.

Code along template

<https://github.com/ajstewartlang/CarpentryConnect2019/blob/master/template.R>

Let's start coding...

```
# Simulate data ####  
# Sample A from population with mean = 1000, sd = 20  
# Sample B from population with mean = 1010, sd = 20  
# This equates to a Cohen's d of .5 as  $(1010-1000)/20$   
# so a medium effect size
```

