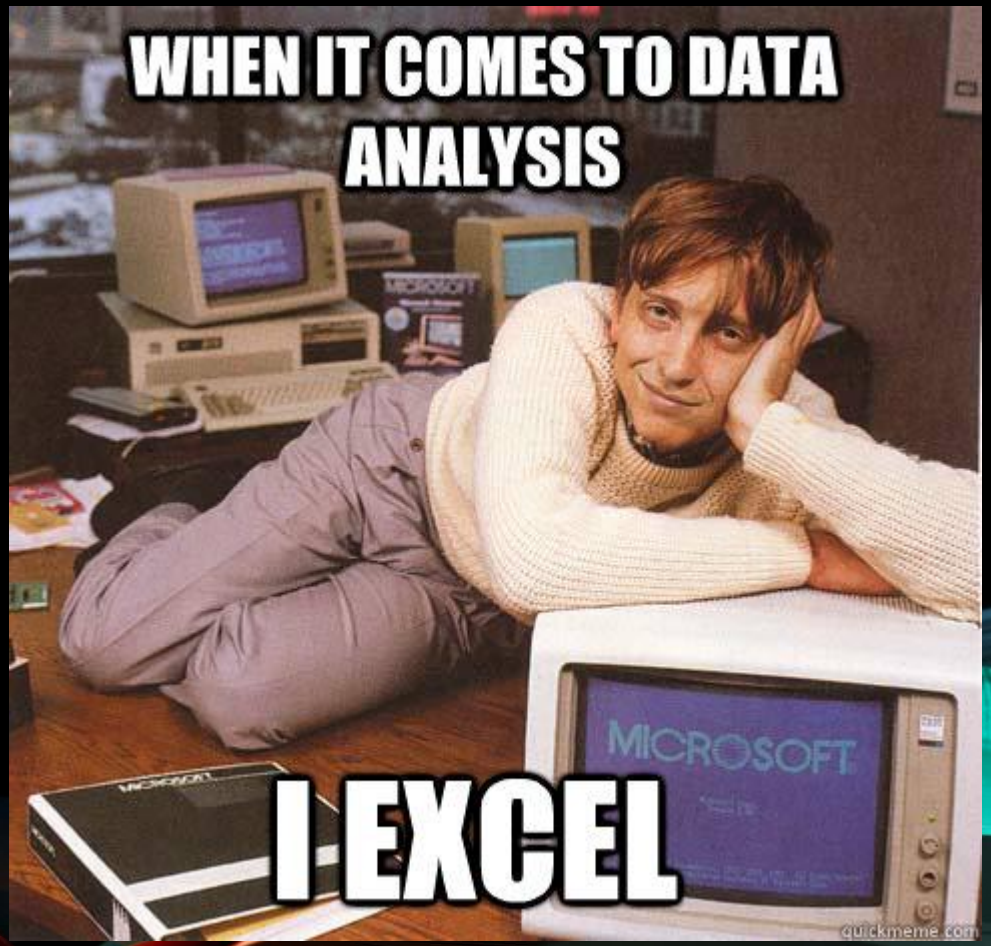
EXPLORATORY DATA ANALYSIS (EDA)



MY BACKGROUND

- BS Chem Eng at CSM 2009
- PhD UCSB 2013
  - Studied phosphor materials in LED lighting
- Worked at thin-film solar cell factory for about 1y (NuvoSun)
- Tried building water-cooled LED lighting
- Went to Galvanize DSI

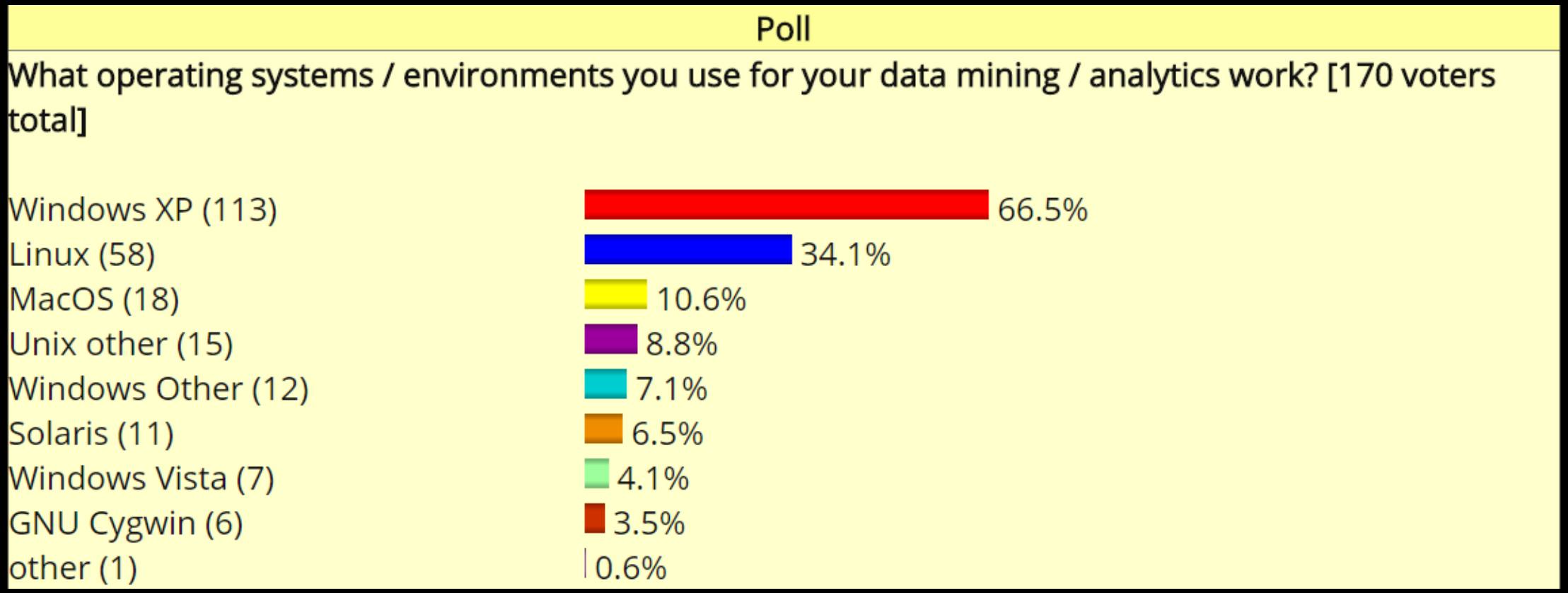
SYLLABUS

- Office hours: by appointment (423 in Clark Hall)

Assignments	Weighted Percentage
Discussion Questions (8 at 1.25% each)	10%
Assignments (13 Assignments at 6.925% each)	90%
Total	**100%**

FIRST SOME NOTES ON OPERATING SYSTEMS

- Anyone know of/read kdnuggets.com ?
- Poll: http://www.kdnuggets.com/polls/2007/operating_systems_for_data_mining.htm



FIRST SOME NOTES ON OPERATING SYSTEMS

- Although people use Windows for analytics, this is probably due to some software (*cough cough* Tableau) that only runs on Windows and OSX
- For programming/general computing, Linux/Unix is typically best
  - OSX is based on UNIX
  - Linux Mint MATE, Ubuntu MATE
  - Linux has far less bloatware and OS junk that Windows has, meaning everything runs faster than in Windows
  - The terminal is much better in Linux/Unix than in Windows
- I recommend dual-booting. Instructions here
  - Backup any important files before this just in case!
  - You can also run a virtual machine, but if your computer is slow, this isn't an option

FIRST SOME NOTES ON OPERATING SYSTEMS

- My setup:
- Ubuntu Linux 16.04, dual booting Windows 10
  - I pretty much always use Ubuntu (the latest LTS version), use virtualbox for Windows (Tableau, adobe acrobat pro)
- HP Laptop, about $600-$700 in mid 2014 (Costco)

TRICKS FOR FASTER WORKFLOW

- Tab completion
- ctrl+arrow left and right will move around a word at a time

INSTALL R KERNEL AND TRY SOME JUPYTER NOTEBOOK STUFF