~~Intro to Bioinformatics – Handling Data~~

Tidy Data
(and why you care!?)

C. Titus Brown

Assoc Prof

VM: Population Health and Reproduction

May 7, 2016

A little exercise, first!

- Could everyone please write down on notecard (non-students too!):
  - A favorite date, preferably one near your birthday;
  - A favorite geographical location (city, etc.), e.g. a place near where you grew up?

Who am I?

- New prof in SVM, working on genetics and genomics (horse, dog, etc.)
  - (not a vet!)
- Focused on “Big Data” problem.
- Volunteered to organize a half-day workshop for y’all on Tidy Data.

- Blown away by study abroad and STAR proposals!!!

Data entry/analysis can be disastrous!

- Date conversion
- Missing year/etc problems
- European formatting
- Doing date conversion is easy for ~20… but for 100s or 1000s?

Now that you know, you won't fall into *this* trap... But there are plenty of other traps!

”Tidy data”

“How can I coordinate data gathering and entry so that I and other can ask precise questions of my data?”

** note: “others” means “you in 6 months”

- General principles of data organization.
- Tools to help you avoid making mistakes.
- A few tips and tricks.
- Ready translation to large(r)-scale analysis (R, Python) & some basic demos.

Why????

- Data entry and analysis is super important in research and clinic.
- Lots of data coming & volume growing… clinical, genetic, sensor, health records, Internet, database...

Prediction: one of your big challenges in research and/or clinic will be (is?) in *finding* things relevant to today's question… think about a system, and keep thinking!

(How do you all organize yourselves now? How do you find e-mail? Will it scale to 100s of messages a week, or a day?)

How could we have done data entry better?

- Dates… ??

- Locations – what was our (possible) goal?

What will the workshop actually be about?

We'll be poking at data, showing you some tools to help you deal with it, and being as "fun" as I can make something mostly computational ;)

- General principles of data organization.
- Tools to help you avoid making mistakes.
- A few tips and tricks.
- Ready translation to large(r)-scale analysis (R, Python) & some basic demos.
- **Using The Google**