Bayes class 6 – Name:

Sean van der Merwe

# Issues with real data

In this class we are going to look at issues that come up when working with real data.

# Some issues

Prior to class each person was assigned one of the following issues to look up and explain to the class. We will go through them now in the order below.

NB: There are lots of other issues not covered here, this is just to get the conversation going.

For each issue, listen to what the other students are saying about it and then write down some key terms or phrases that catch your attention or interest. Look up the things you find interesting after class, to broaden your understanding.

## Representativeness of samples

## Selection issues in sampling

## Missing data

## Censored data

## Typos and inconsistencies

## Outliers

## Severe skewness and extreme values

## Confounding variables

## Multicollinearity

## Multiple testing

## Data arrangements (e.g. long or wide)

## Unbalanced samples

## Data transformations

## Other things mentioned that might be worth reading up on