In-person session 2

January 20, 2022

PMAP 8521: Program evaluation
Andrew Young School of Policy Studies

Plan for today

Regression FAQs

Miscellaneous R stuff

Visualizing data with **ggplot2**

Transforming data with **dplyr**

Regression FAQs

Histogram bin widths

Density plots

Why use two steps to create a regression in R? (i.e. assigning it to an object with `<-`?)

Why use `tidy()` from the broom package?

How was the 0.05 significance threshold determined?

Could we say something is significant if p > 0.05, but just note that it is at a higher p-value?
Or does it have to fall under 0.05?

Why all this convoluted logic of null worlds?

Different "dialects" of statistics

Frequentist

$P(\text{data} \mid H_0)$

"Regular" statistics; what you've learned (and are learning here)

Bayesian

$P(H \mid \text{data})$

Requires lots of computational power

**Do we care about the actual coefficients or just whether or not they're significant?**

How does significance relate to causation?

If we can't use statistics to assert causation how are we going to use this information in program evaluation?
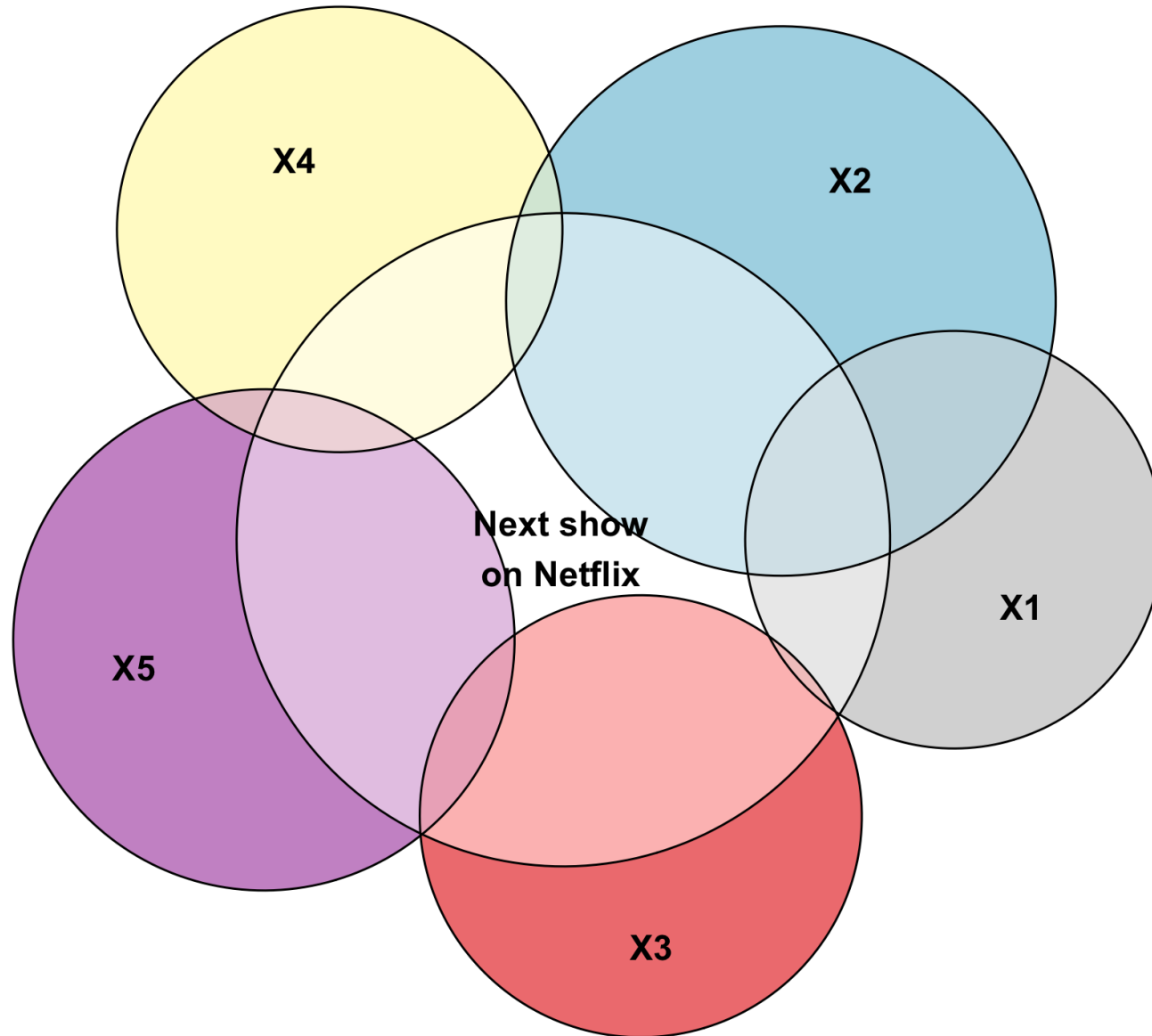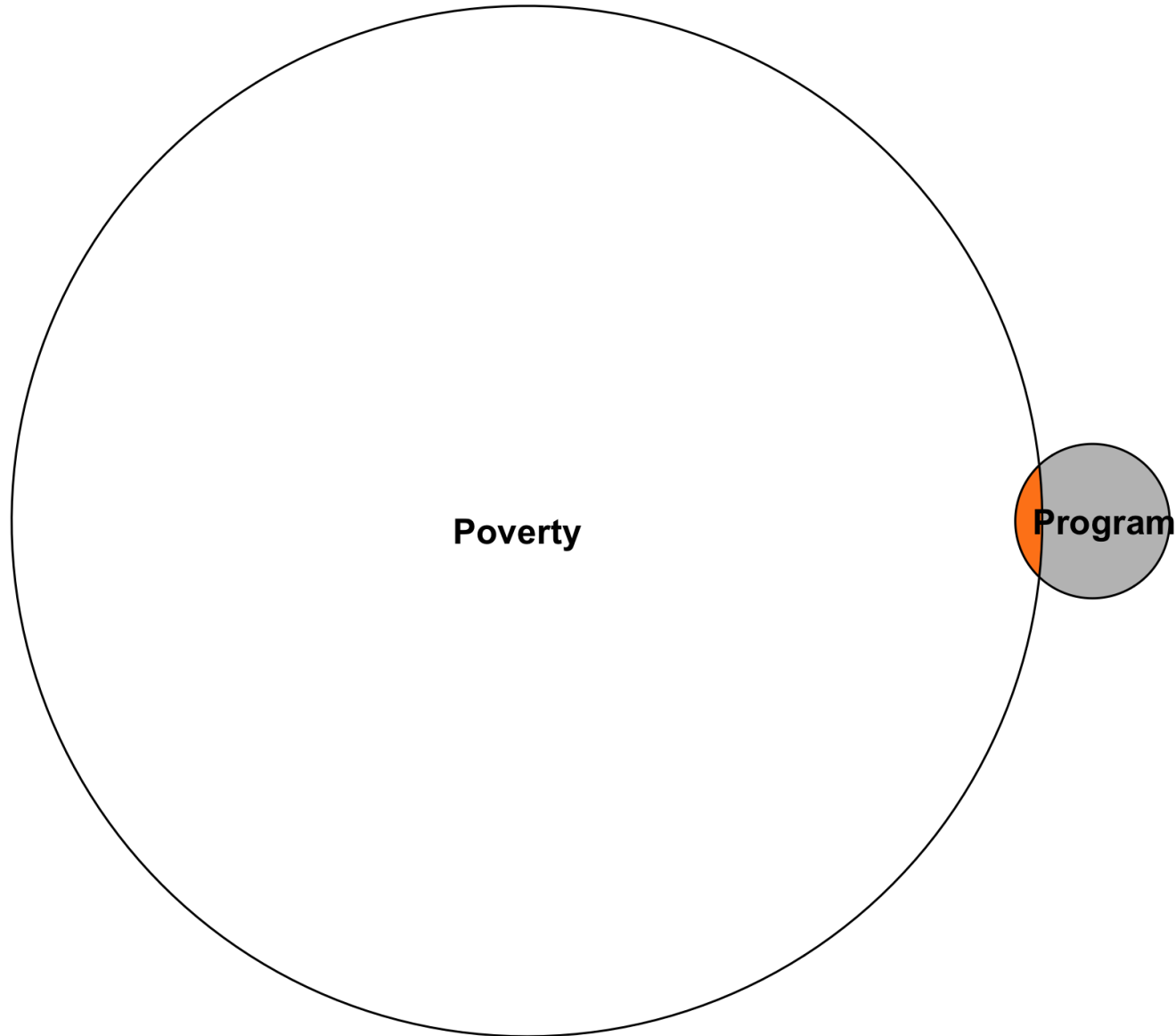
What counts as a "good" R^2?

R² represented as an Euler diagram

Orange area (D + E + G) shows the total variance in outcome Y that is jointly explained by X1 and X2

$$R^2 = \frac{D + E + G}{A + D + E + G}$$

X2
C

E

F

G

Y
A

D

X1
B
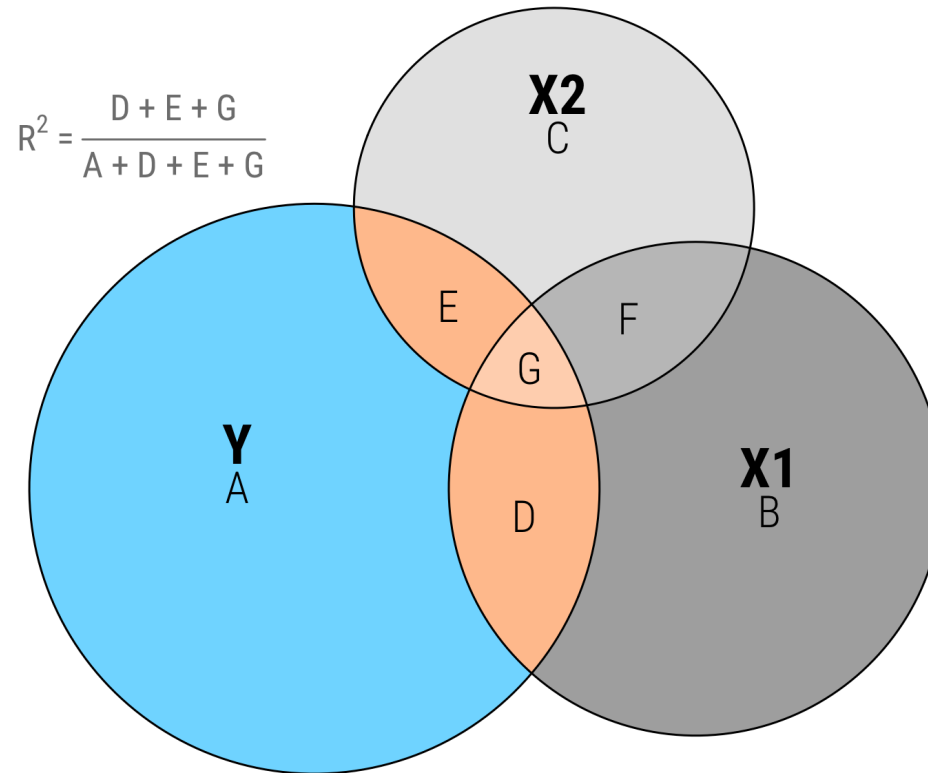
Circles sized according to each variable's sum of squares; size of overlapping areas is not 100% correct due to limitations in available geometric space

Why do we log things?

Miscellaneous R stuff

Miscellaneous R stuff

File paths, working directories, and RStudio projects

Markdown things

Visualizing data with ggplot2

Transforming data with dplyr