

# Introduction to Social Data Analytics

## Class 14

Arushi Kaushik

arkaushi@ucsd.edu

May 2, 2019

# Midterm Format

- One question using Excel, one using Stata
- Question format similar to those from the problem sets
- Have until end of class - please come a little early.

# Topics to be covered: general data concepts

- Sample size, observations, variables
- Sample vs population & representativeness
- Unit of analysis
- Sampling and response biases
- Variable types (numerical, (un)ordered categorical, logical)
- Logical operators (and, or, not)
- Regression

# Topics to be covered: Excel functions and operations

- Insert/delete/sort columns, AutoFill
- Cell references
- Logical (AND, OR, NOT)
- Statistical (COUNT, SUM, AVERAGE, MEDIAN, MAX, MIN)
- Conditional functions (IF, COUNTIF, AVERAGEIF, etc.)
- Search functions (MATCH, VLOOKUP)
- Insert chart (plotting)

# Topics to be covered: Stata functions and operations

- Logical operators (&, |, !)
- Loading and exploring data (e.g. import, browse, count, describe, list, summarize, tabulate)
- Variables (e.g. drop, generate, rename, replace)
- Sorting and selecting (e.g. by, sort, if, in)
- Graphing (e.g. graph, twoway, scatter, line, histogram, bar, box)
- Regression (e.g. regress, lfit)
- Wrangling (e.g. append, merge, collapse, reshape)
- = vs == and " " vs .

# Midterm Survey

[goo.gl/6fPgvY](https://goo.gl/6fPgvY)

Once you've finished, we can use the rest of the time for Q & A.