Rutgers Linguistics Workshop on Mixed Effects Models
— Visualizing data with ggplot2 —

09/17/2016
Judith Degen
Stanford University

What is ggplot?

- an R package for producing statistical graphs

- a language created by Hadley Wickham, based on "The Grammar of Graphics" by Wilkinson et al (2005), for describing and creating plots

- plots can be built up iteratively and updated later

The grammar of graphics

What is a graphic?
A mapping from **data** to **aesthetic attributes** (e.g., color, shape, size) of **geometric objects** (e.g., points, lines, bars)

A graphic
- may contain statistical **transformations** of the data
- is drawn on a specific **coordinate system**
- can use **faceting** for generating the same plot for different subsets of the dataset

Components of a graphic

- **data** and set of **aesthetic mappings** describing how to map variables in data to aesthetic attributes

- geometric objects (**geoms**) are what you see: points, lines, bars

- statistical transformations (**stats**) summarise data (e.g., binning, counting observations) → optional

- **scales** map values in data space to values in aesthetic space

- **coordinate system** describes how data coordinates are mapped to graphic plane

- **faceting** specification describes how to break data into subsets andreturn an individual plot for each subset (like conditioning)