GEOG 5680
Introduction to R
07: Plotting with **ggplot2**

Simon Brewer

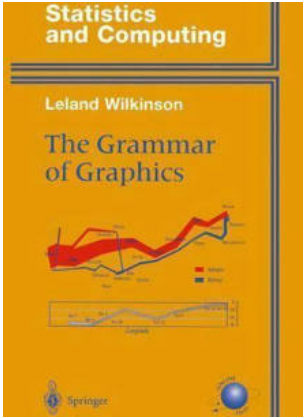
Geography Department
University of Utah
Salt Lake City, Utah 84112

simon.brewer@geog.utah.edu

May 03, 2020

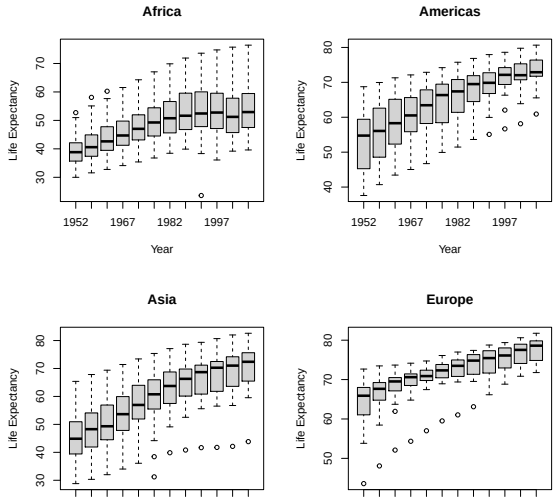
ggplot2

- Based on Leland Wilkinson's Grammar of Graphics
  - All data figures can be represented by the same *grammar*
- Adapted for R by Hadley Wickham
- Provides much easier methods for comparative plots

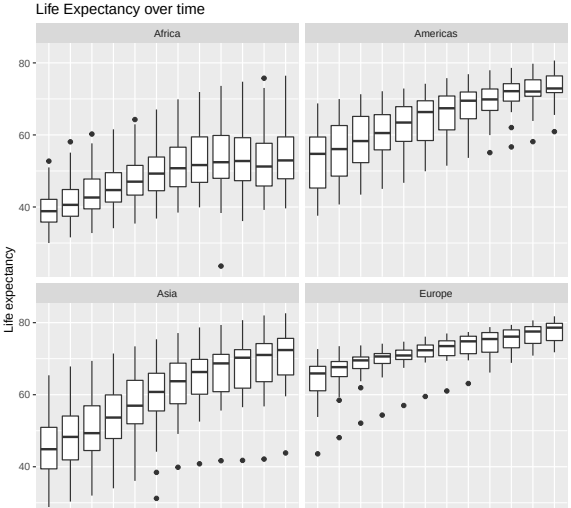


Base graphics vs. ggplot2

Base graphics: 8 lines

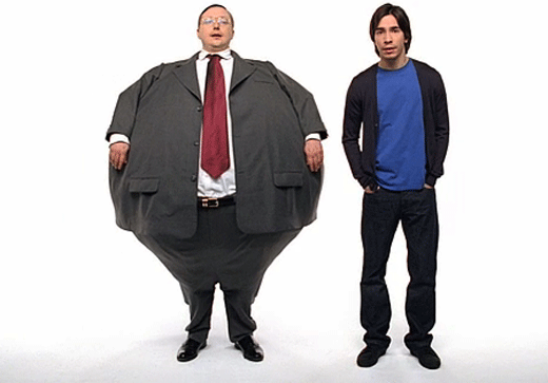


ggplot2: 1 (quite complex) line



Data frames for ggplot2

- Data is often presented as short and fat tables
- Plotting is easier with tall and thin data frames
  - Each variable forms a column
  - Each observation forms a row



Data frames for ggplot2

Short/fat table: good for presenting results

	TreatA	TreatB
Jane Smith	-	2
John Doe	16	11
Mary Jones	3	1

Tall/thin dataframe: preferred for plotting

Name	Treat	Result
Jane Smith	a	-
John Doe	a	16
Mary Jones	a	3
Jane Smith	b	2
John Doe	b	11
Mary Jones	b	1

- Support package **reshape2** includes functions to transform between these layouts
- `cast`: thin data frame to table
- `melt`: table to thin data frame

Grammar of Graphics

- Theme: font, colors, etc
- Coordinates: plotting space
- Statistics: summaries/transformation of data
- Facet: subplots based on groups
- Geometry: shapes used to represent data
- Aesthetic: scale used to map data
- Data: as data frame