Tips for Effective Data Visualization

Angela Zoss · Eric Monson
Data and Visualization Services
STA 199L · Spring 2018

Slides: http://bit.ly/STA199LVisSpring2018
1 simple dataset: Which is the best chart? And why?
Even with just 20 values, this is a difficult question to answer.

Andy Cotgreave, Tableau
@acotgreave

https://www.youtube.com/watch?v=AuJFuEq-qD8
ggplot2
Principles for Effective Visualizations
Principle 1: Order matters
How much experience do you have as a producer (e.g., reader, follower) of network science research?
data$answer <-
factor(data$answer,
  levels=c("None", "A little", "Some", "A lot"),
  ordered = TRUE)
How much experience do you have as a producer (e.g., reader, follower) of network science research?
Order by value

data$academic_field <-
  factor(data$academic_field,
         levels=names(
           sort(
             table(
               data$academic_field), decreasing=TRUE)))
Principle 2: Put long categories on y-axis
Flip the axes

coord_flip()
Oops!

data$academic_field <- factor(data$academic_field, levels=names (sort (sort(table(data$academic_field), decreasing=TRUE))))
Primary academic field

- Sociology: 10.0
- Business: 7.5
- Computer sciences: 9.0
- Communication studies: 8.0
- Skipped: 0.0
- Physics: 2.5
- Economics: 3.0
- Anthropology: 4.0
- Public administration: 5.0
- Psychology: 3.0
- Other: 5.0
- History: 7.5
- Political science: 2.0
- Medicine: 0.0
- Mathematics: 1.0
- Information science: 0.0
- Geography: 0.0
- Engineering: 1.0
Principle 3: Pick a purpose
Different placement helps with different comparisons

\[
\text{fill=\text{highest\_degree}}
\]

\[
\text{facet\_grid(\text{.~highest\_degree})}
\]
Principle 4: Keep scales consistent
Keep all categories, manually set axes

```r
scale_x_discrete(drop=FALSE)
scale_y_continuous(limits=c(0,40),
                  breaks=c(0,10,20,30,40),
                  minor_breaks=NULL)
```
How much experience do you have as a producer of network science research?

How much experience do you have as a consumer of network science research?
Principle 5: Select meaningful colors
Select colors manually, or use alternate palette

```r
scale_fill_manual(
  values=c("snow4","snow3",
           "tan3","tan1",
           "turquoise2","turquoise4"))

scale_fill_manual(
  values=c("#fee391","#fe9929","#cc4c02"))

# Also see package RColorBrewer
scale_fill_brewer(palette="BrBG")
```
ggplot2 Resources

• General ggplot2 information
  http://ggplot2.tidyverse.org/

• R Graphics Cookbook (recipes for plots)
  http://www.cookbook-r.com/Graphs/index.html

• R for Data Science (online book that includes ggplot2)
  http://r4ds.had.co.nz/

• ggplot2: Elegant Graphs for Data Analysis (book by Hadley Wickham)
  http://ggplot2.org/book/

• ggplot2 cheatsheet (also in RStudio)
Resources
Data and Visualization Services

http://library.duke.edu/data
askdata@duke.edu
Information about DVS

• Data collections, LibGuides, etc.
  http://library.duke.edu/data/

• Blog (tutorials, announcements, etc.)
  http://blogs.library.duke.edu/data/

• E-mail consultations
  askdata@duke.edu

• Mailing list for announcements:
  https://lists.duke.edu/sympa/subscribe/dvs-announce

• Twitter accounts
  @duke_data, @duke_vis
Videos of past workshops

Questions?
askdata@duke.edu
Slides: https://duke.box.com/v/IntroDataVisS18
Resources
Data and Visualization Services

http://library.duke.edu/data
askdata@duke.edu
Information about DVS

- Data collections, LibGuides, etc.  
  http://library.duke.edu/data/

- Blog (tutorials, announcements, etc.)  
  http://blogs.library.duke.edu/data/

- E-mail consultations  
  askdata@duke.edu

- Mailing list for announcements:  
  https://lists.duke.edu/sympa/subscribe/dvs-announce

- Twitter accounts  
  @duke_data, @duke_vis
Videos of past workshops

Questions?
askdata@duke.edu
http://bit.ly/STA199LVisSpring2018