Tips for Effective Data Visualization

Angela Zoss · Eric Monson
Data and Visualization Services

STA 199L · Spring 2018

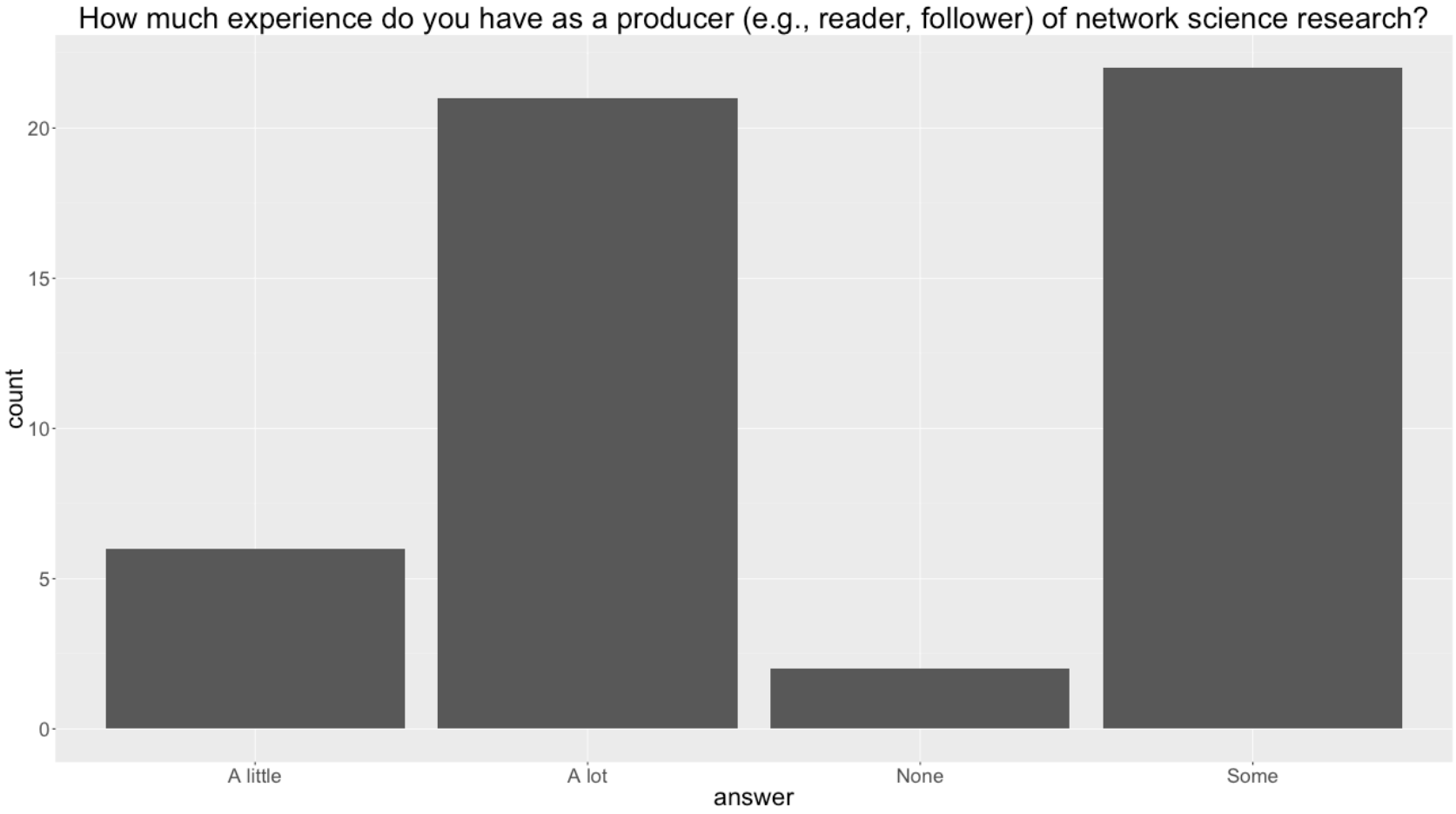
Slides: http://bit.ly/STA199LVisSpring2018

https://www.youtube.com/watch?v=AuJFuEq-qD8

ggplot2

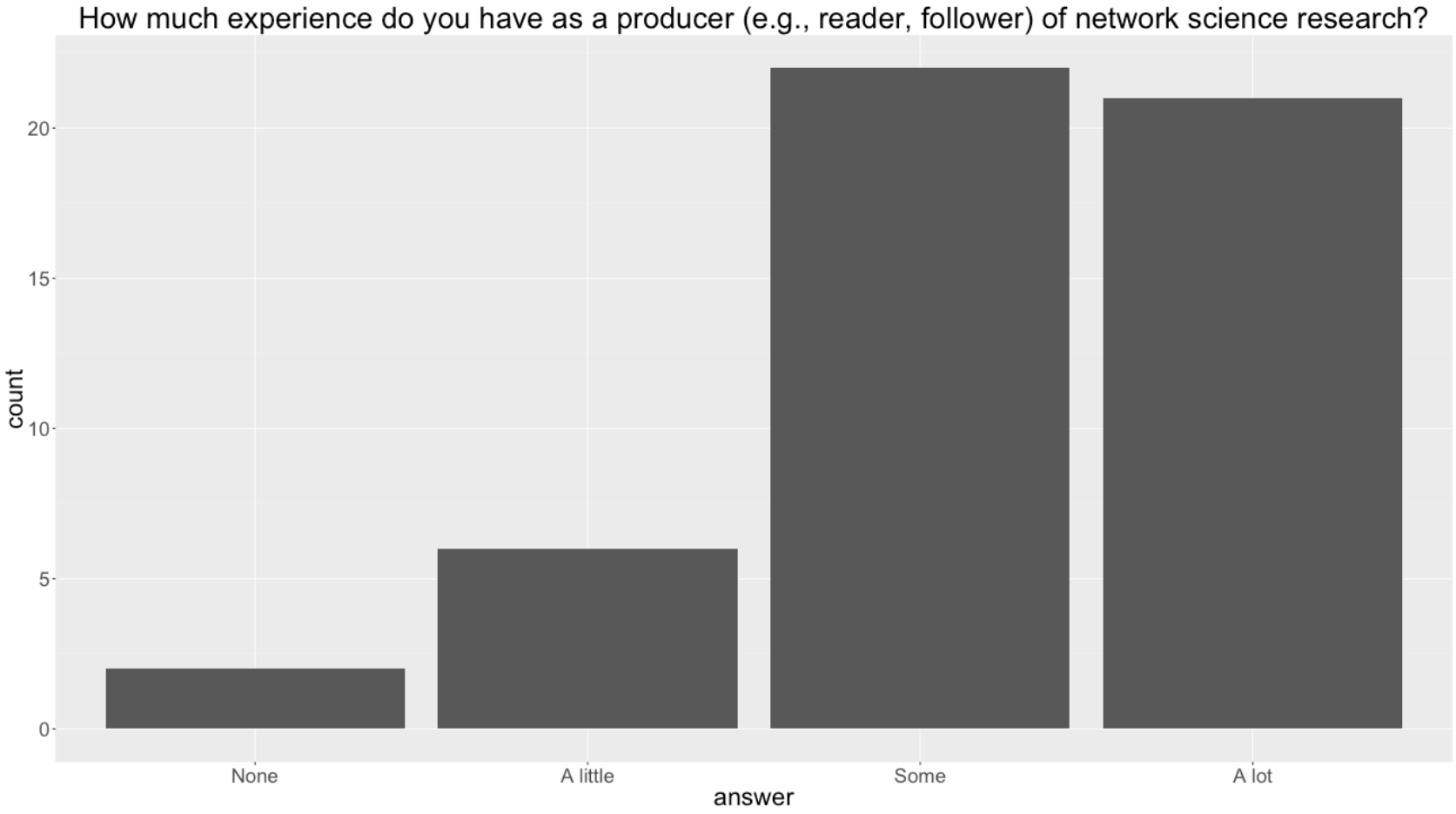
Principles for Effective Visualizations

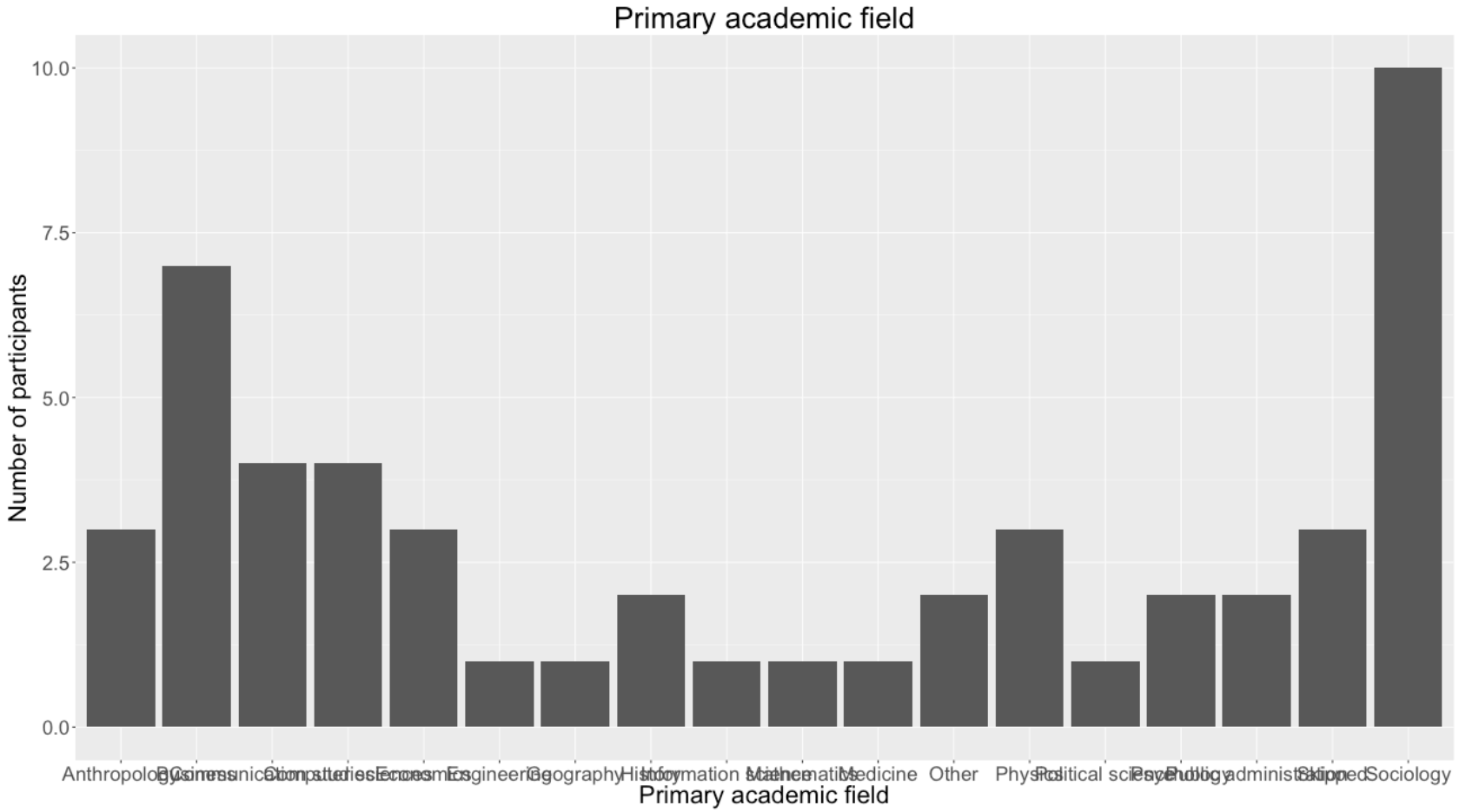
Principle 1: Order matters


How much experience do you have as a producer (e.g., reader, follower) of network science research?
count
20
15
10
5
0
A little
A lot
None
Some
answer

Order by meaning

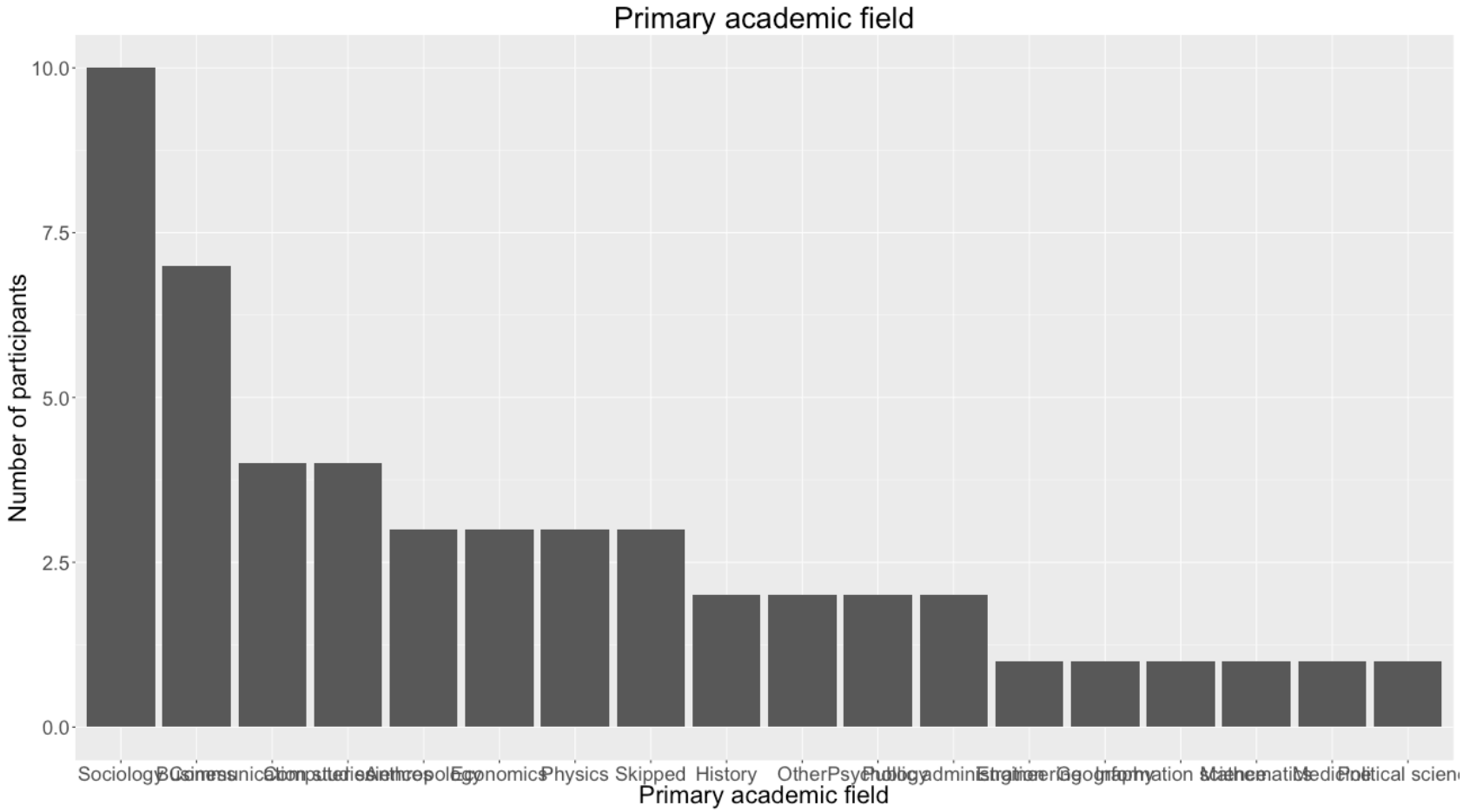
```
data$answer <-
    factor(data$answer,
           levels=c("None", "A little", "Some", "A lot"),
           ordered = TRUE)
```


How much experience do you have as a producer (e.g., reader, follower) of network science research?
count
20
15
10
5
0
None
A little
Some
A lot
answer

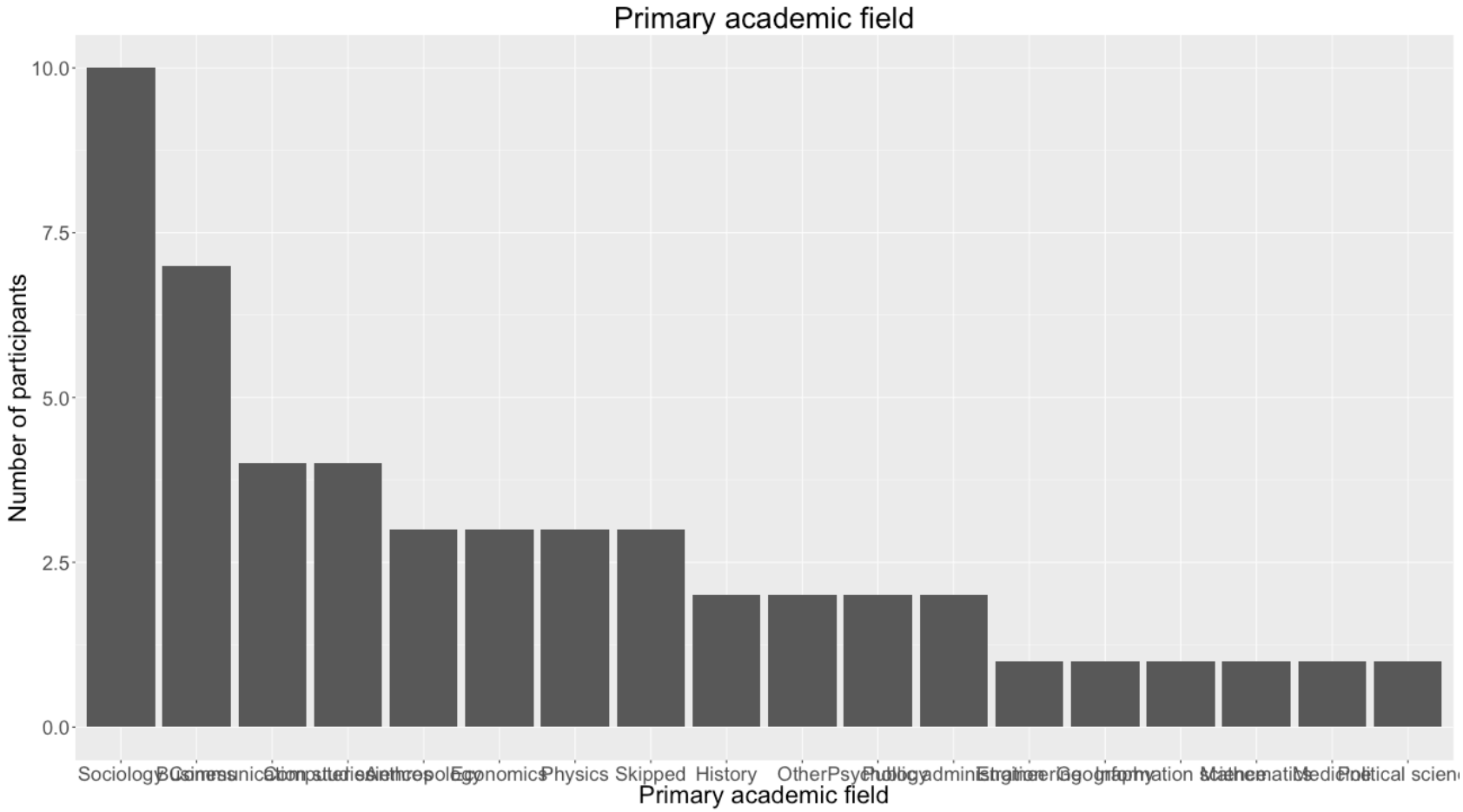

Primary academic field
Number of participants
10.0
7.5
5.0
2.5
0.0
Anthropology
Business
Communication
Computer science
Economics
Engineering
Geography
History
Information science
Mathematics
Medicine
Other
Physics
Political science
Psychology
Public administration
Sociology
Primary academic field

Order by value

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

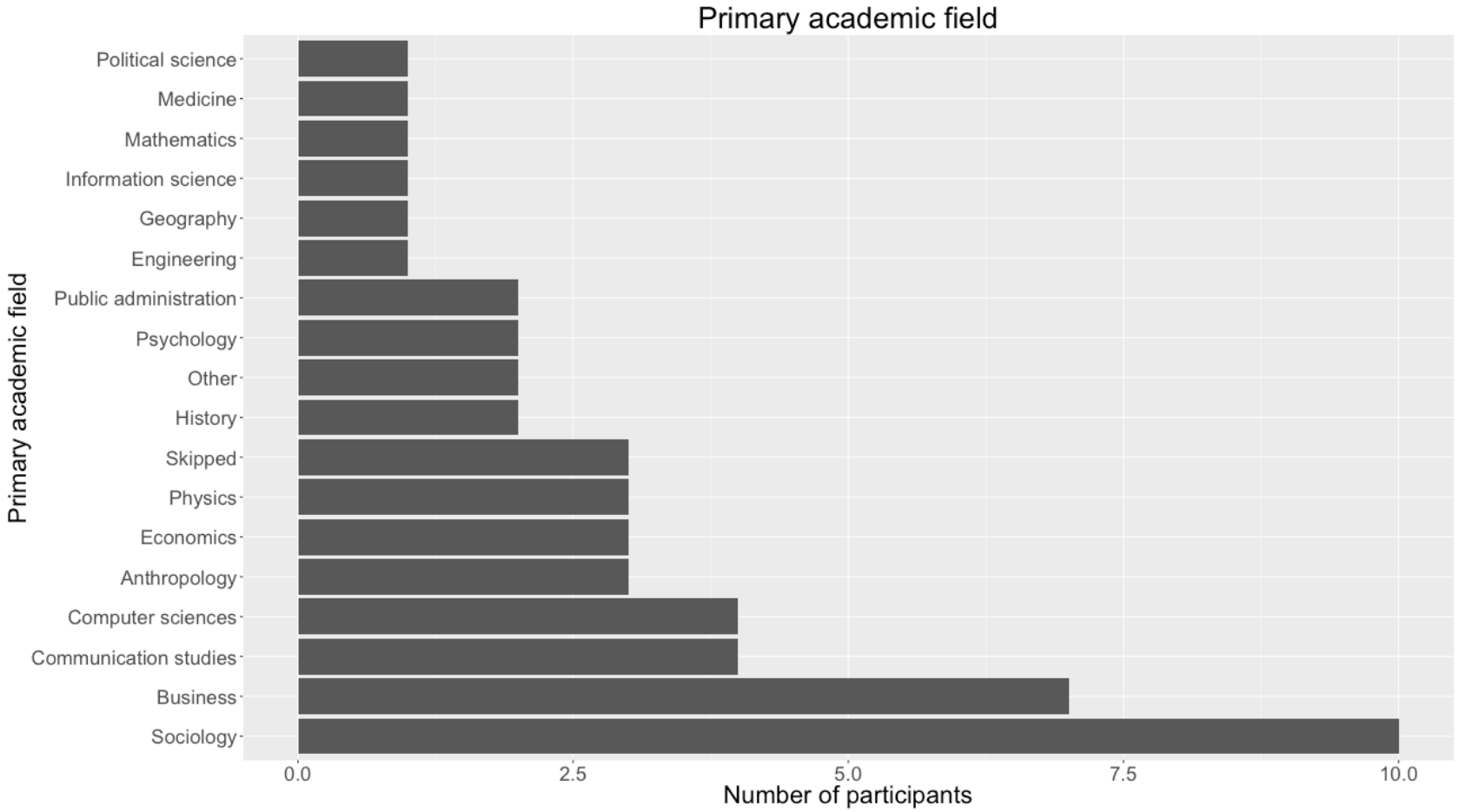

Primary academic field
Number of participants
10.0
7.5
5.0
2.5
0.0
Sociology
Business
Communication
Computer science
Anthropology
Economics
Physics
Skipped
History
Other
Psychology
Public administration
Engineering
Geography
Information science
Mathematics
Medicine
Political scien
Primary academic field

Principle 2: Put long categories on y-axis


Primary academic field
Number of participants
10.0
7.5
5.0
2.5
0.0
Sociology
Business
Communication
Computer science
Anthropology
Economics
Physics
Skipped
History
Other
Psychology
Public administration
Engineering
Geography
Information science
Mathematics
Medicine
Political scien
Primary academic field

Flip the axes

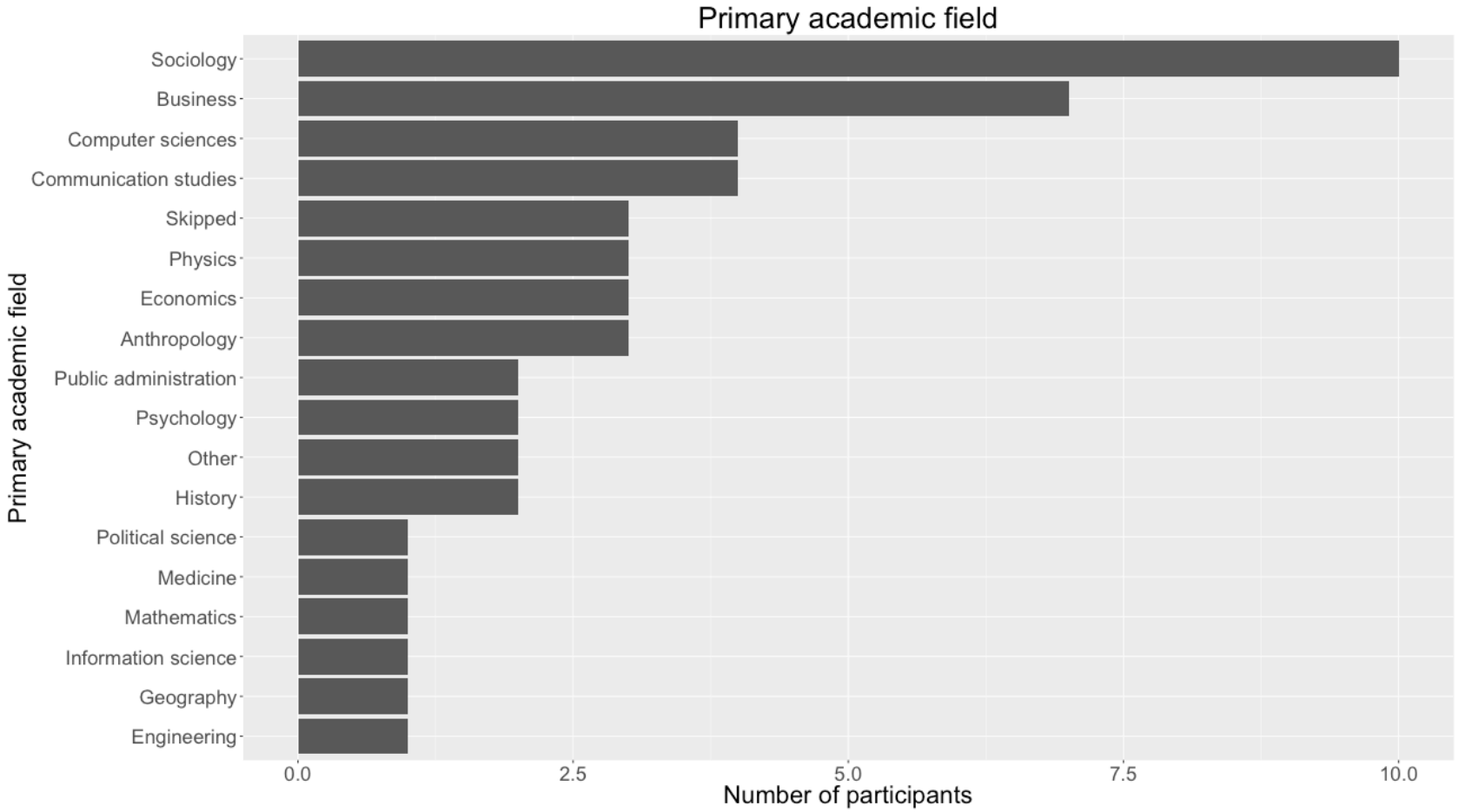
```
coord_flip()
```


Primary academic field
Political science
Medicine
Mathematics
Information science
Geography
Engineering
Public administration
Psychology
Other
History
Skipped
Physics
Economics
Anthropology
Computer sciences
Communication studies
Business
Sociology
Primary academic field
0.0
2.5
5.0
7.5
10.0
Number of participants

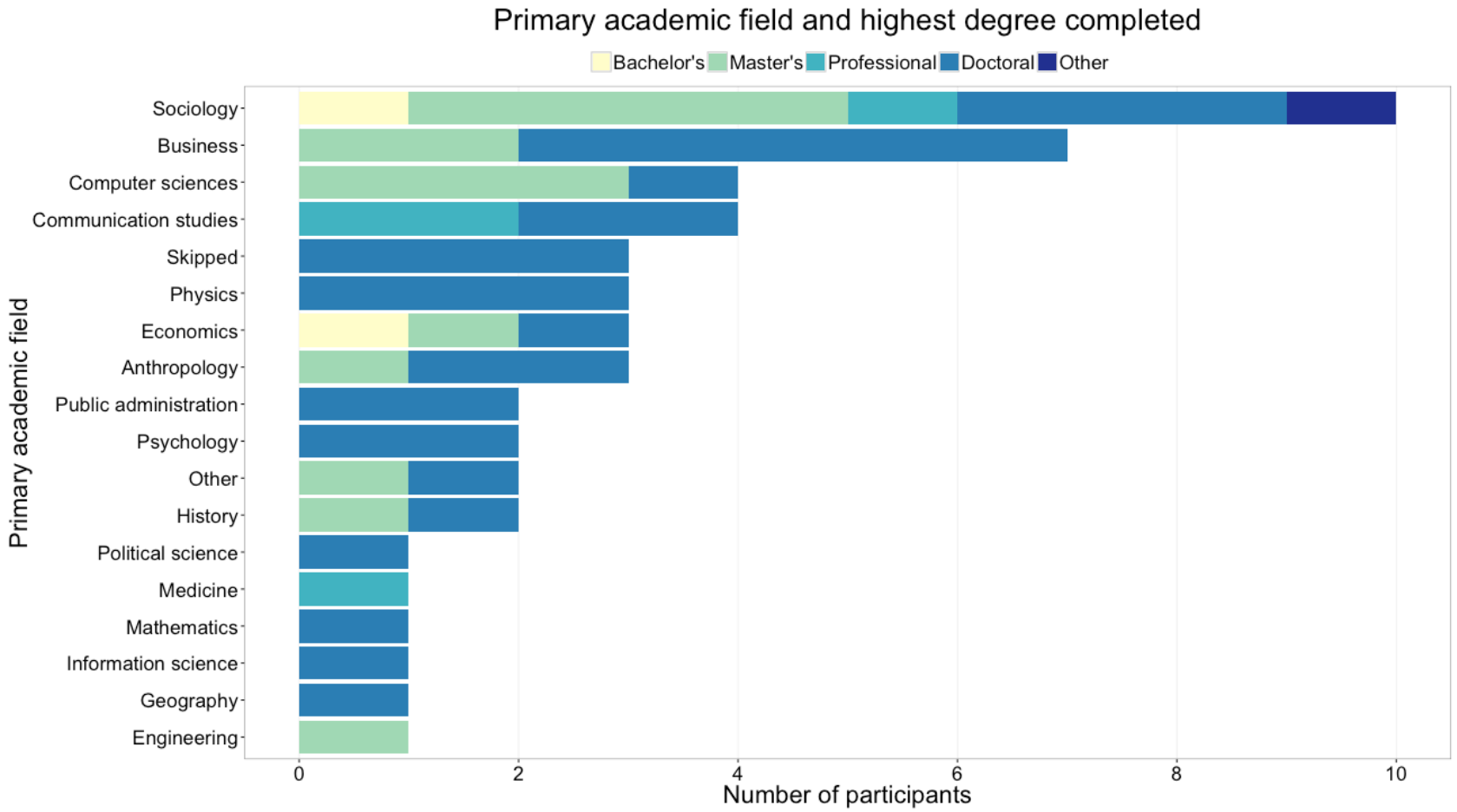
Oops!

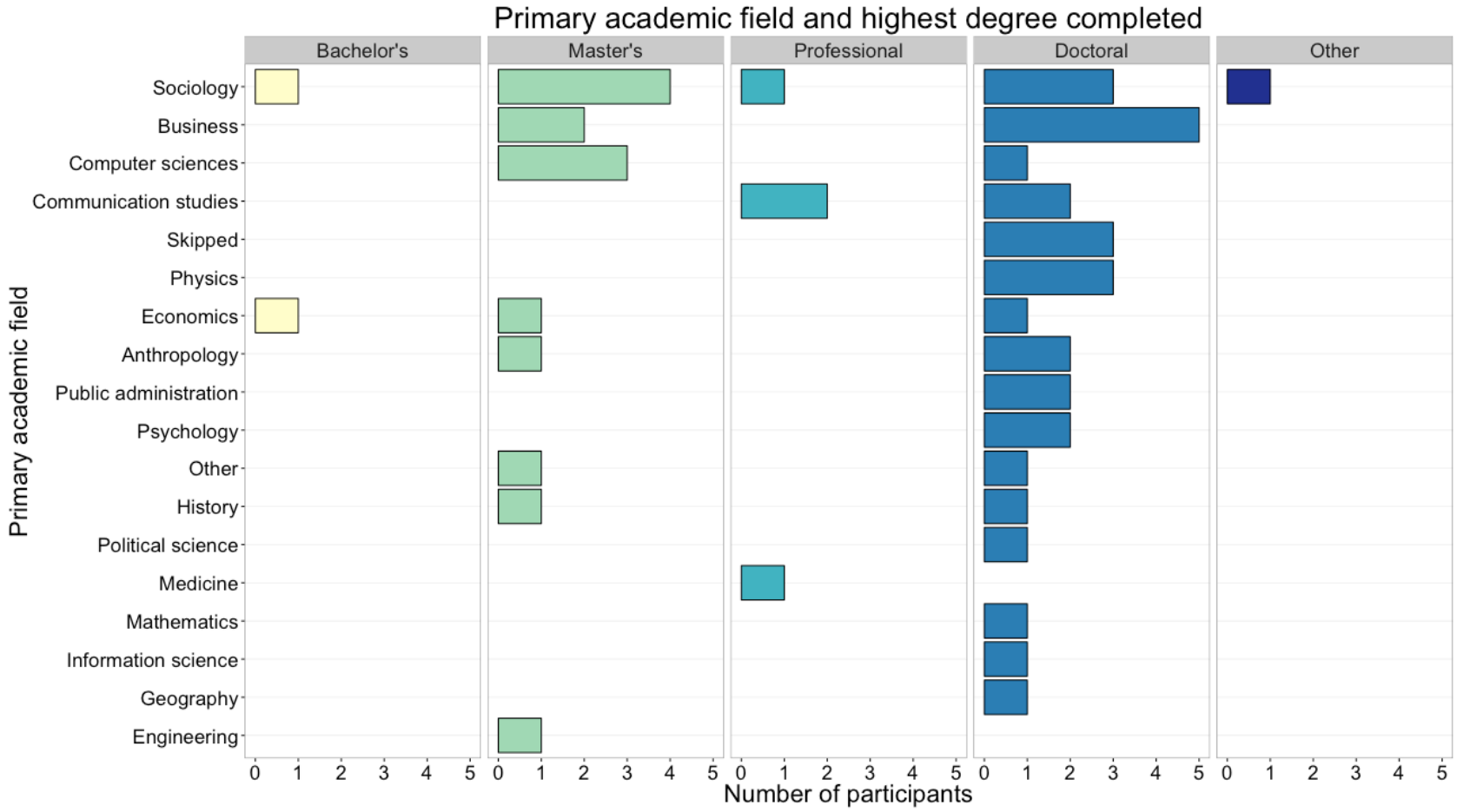
```
data$academic_field <-
     factor(data$academic_field,
           levels=names(
               sort(
                   table(data$academic_field),
                   decreasing=TRUE)))
```

```
data$academic_field <-
     factor(data$academic_field,
           levels=names(
               sort(
                   table(data$academic_field))))
```


Primary academic field
Sociology
Business
Computer sciences
Communication studies
Skipped
Physics
Economics
Anthropology
Public administration
Psychology
Other
History
Political science
Medicine
Mathematics
Information science
Geography
Engineering
Primary academic field
0.0
2.5
5.0
7.5
10.0
Number of participants

Principle 3: Pick a purpose


Primary academic field and highest degree completed
Bachelor's
Master's
Professional
Doctoral
Other
Primary academic field
Sociology
Business
Computer sciences
Communication studies
Skipped
Physics
Economics
Anthropology
Public administration
Psychology
Other
History
Political science
Medicine
Mathematics
Information science
Geography
Engineering
0
2
4
6
8
10
Number of participants

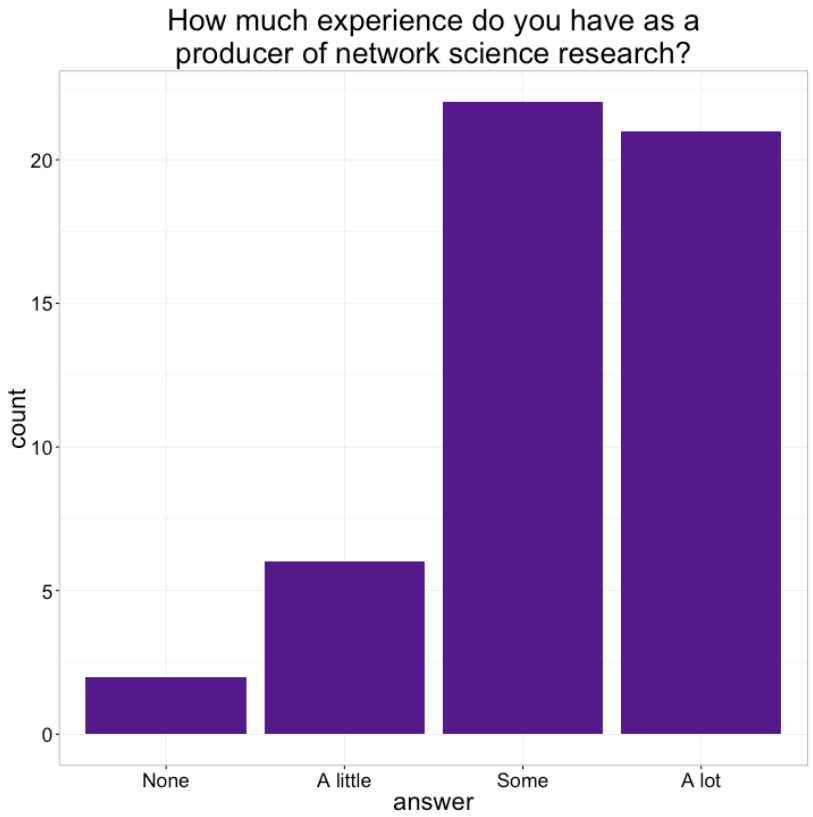

Primary academic field and highest degree completed
Bachelor's
Master's
Professional
Doctoral
Other
Primary academic field
Sociology
Business
Computer sciences
Communication studies
Skipped
Physics
Economics
Anthropology
Public administration
Psychology
Other
History
Political science
Medicine
Mathematics
Information science
Geography
Engineering
0 1 2 3 4 5
Number of participants

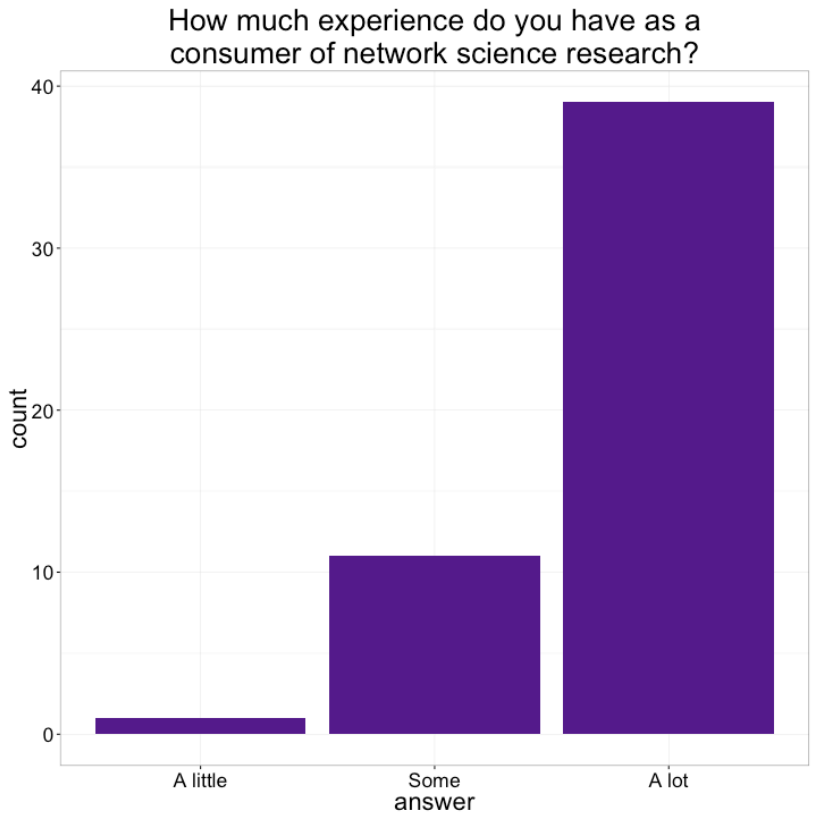
Different placement helps with different comparisons

```
fill=highest_degree
```

```
facet_grid(.~highest_degree)
```

Principle 4: Keep scales consistent

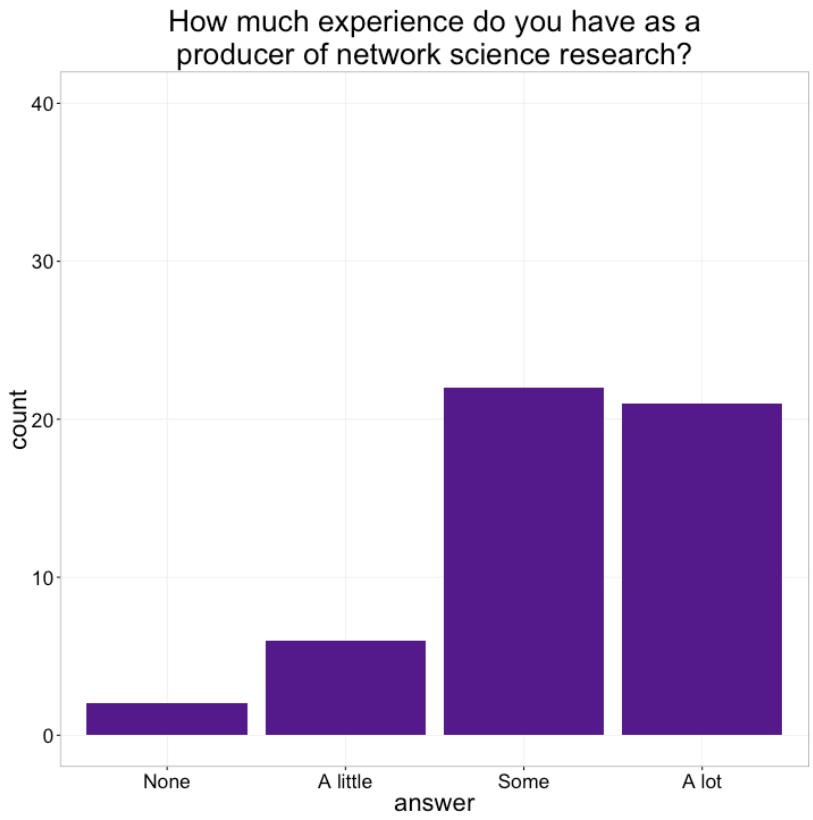

How much experience do you have as a producer of network science research?
count
0
5
10
15
20
None
A little
Some
A lot
answer

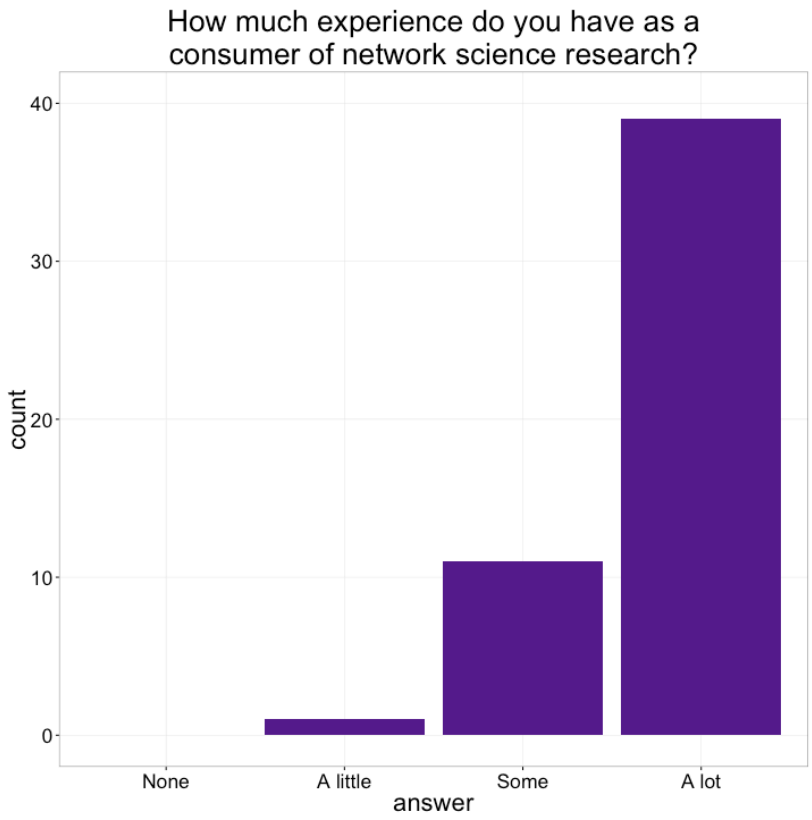

How much experience do you have as a consumer of network science research?
count
0
10
20
30
40
A little
Some
A lot
answer

Keep all categories, manually set axes

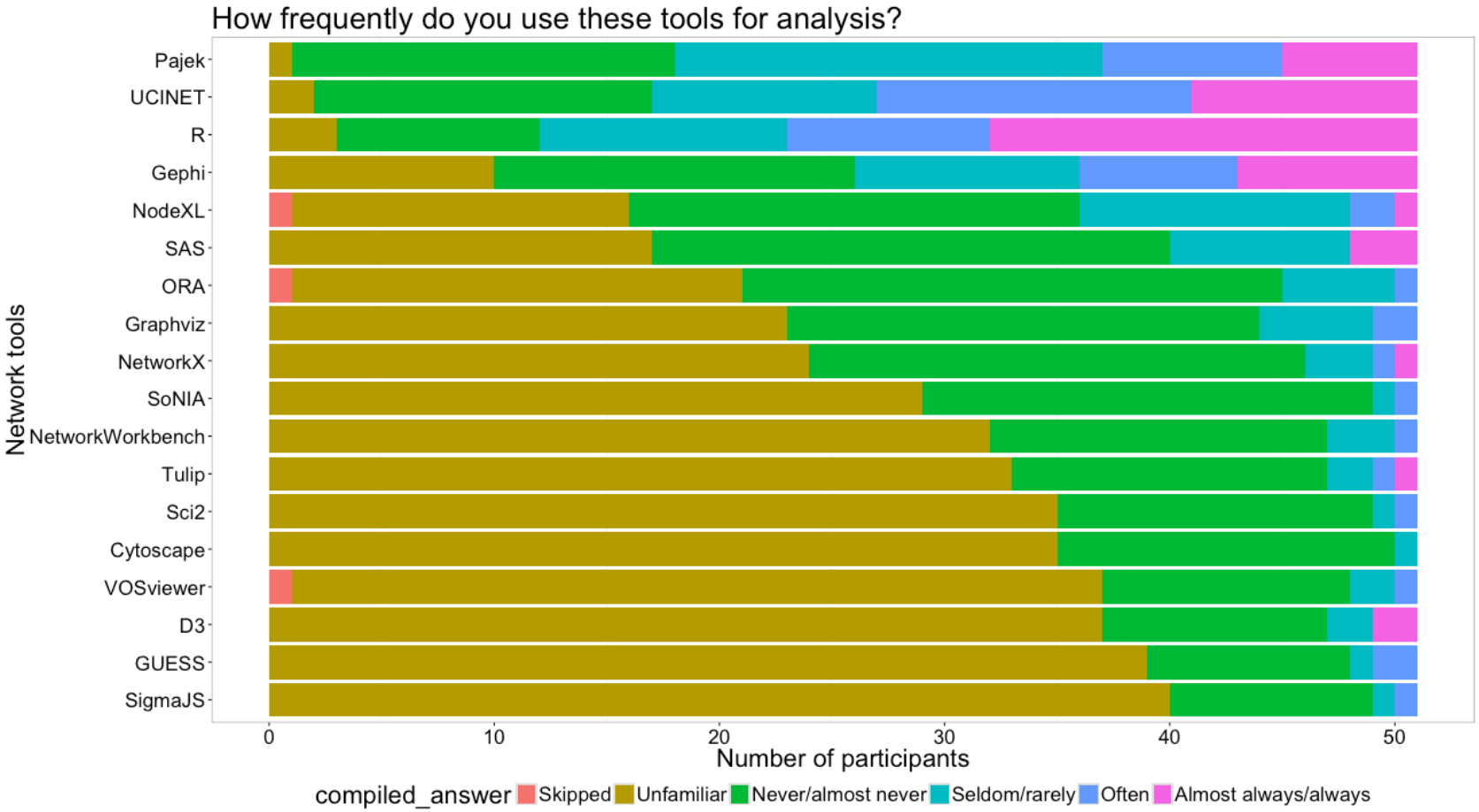
```
scale_x_discrete(drop=FALSE)

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?
count
40
30
20
10
0
None
A little
Some
A lot
answer


How much experience do you have as a consumer of network science research?
count
40
30
20
10
0
None
A little
Some
A lot
answer

Principle 5: Select meaningful colors

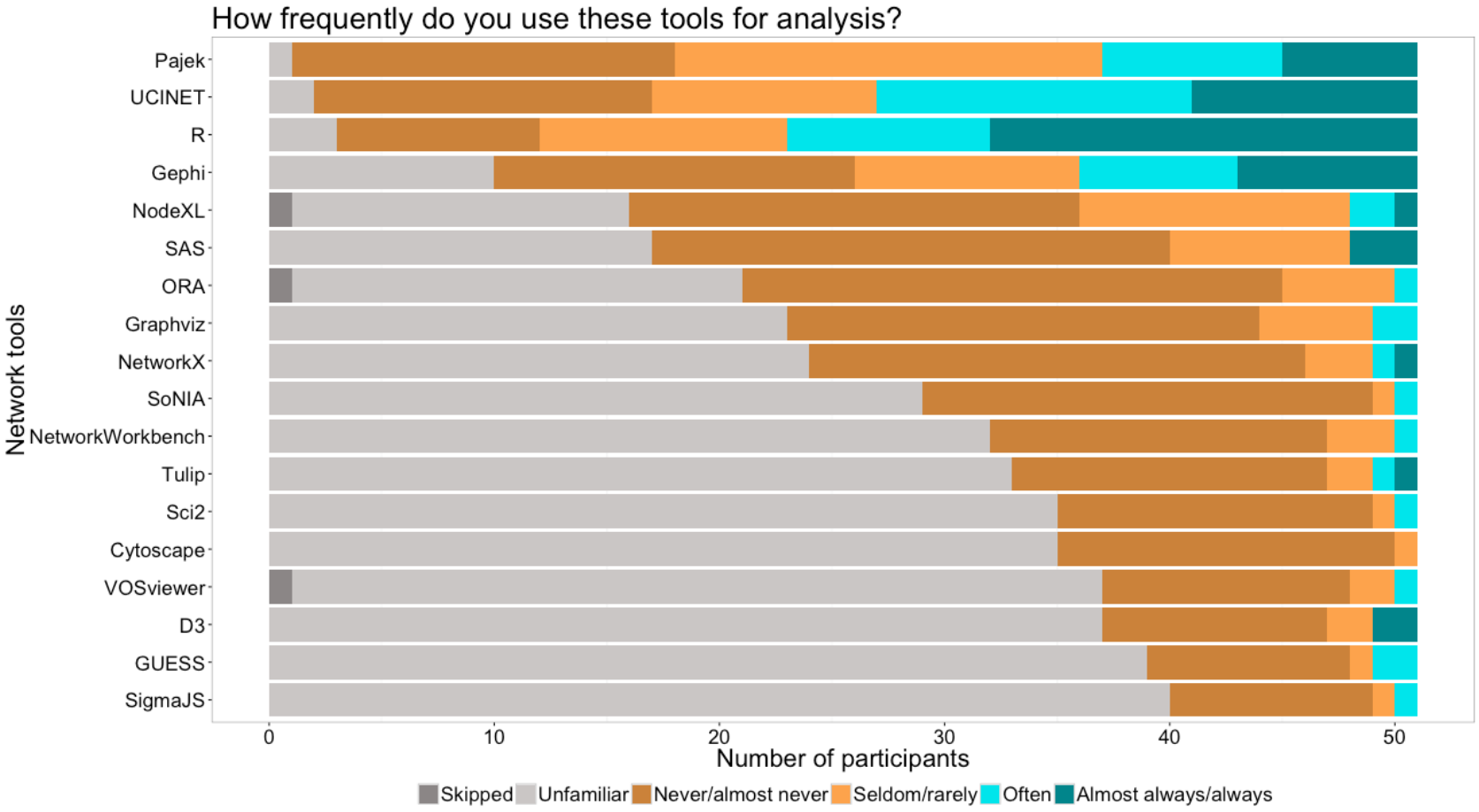

How frequently do you use these tools for analysis?
Network tools
Pajek
UCINET
R
Gephi
NodeXL
SAS
ORA
Graphviz
NetworkX
SoNIA
NetworkWorkbench
Tulip
Sci2
Cytoscape
VOSviewer
D3
GUESS
SigmaJS
0
10
20
30
40
50
Number of participants
compiled_answer
Skipped
Unfamiliar
Never/almost never
Seldom/rarely
Often
Almost always/always

Select colors manually, or use alternate palette

```
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")
```


How frequently do you use these tools for analysis?
Network tools
Pajek
UCINET
R
Gephi
NodeXL
SAS
ORA
Graphviz
NetworkX
SoNIA
NetworkWorkbench
Tulip
Sci2
Cytoscape
VOSviewer
D3
GUESS
SigmaJS
0
10
20
30
40
50
Number of participants
Skipped
Unfamiliar
Never/almost never
Seldom/rarely
Often
Almost always/always

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-book.org/
- ggplot2 cheatsheet (also in RStudio)

https://rstudio.github.io/cheatsheets/data-visualization.pdf

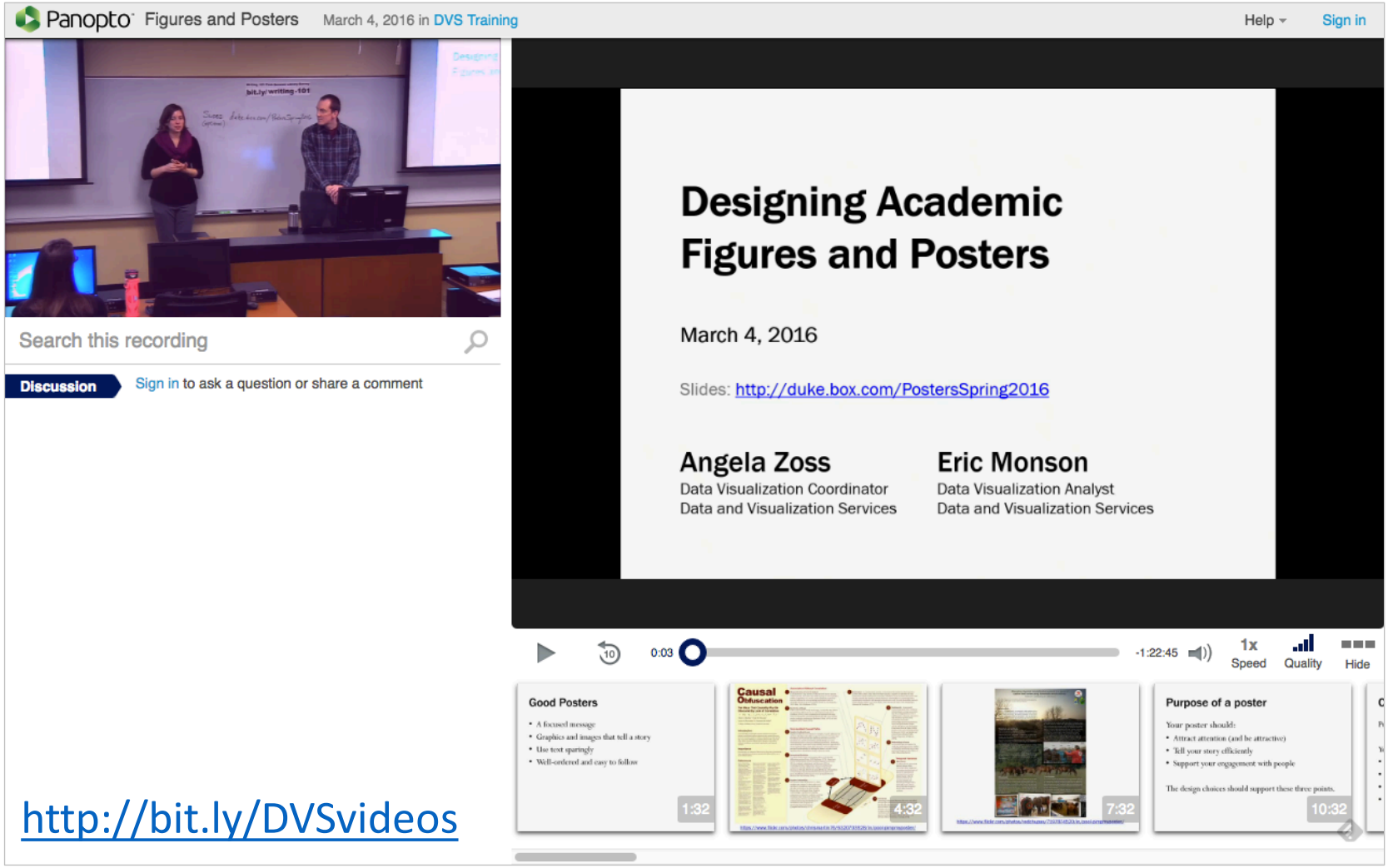
Resources

Data and Visualization Services

Data and Visualization Services Department

http://library.duke.edu/data
askdata@duke.edu

Videos of past workshops



http://bit.ly/DVSvideos