Module B: Knowing Your Data

Slide Deck B1:

Graphics for Categorical Variables

The section in which we see some typical graphics used to understand/see categorical variables. Remember that the researcher must understand their data before trying to analyze it. Graphics help with that.

Today's Objectives

By the end of this slidedeck, you should

- create the following univariate graphics for categorical variables:
  - pie chart
  - bar chart
- create the following *bi*variate graphics for categorical variables:
  - side-by-side bar chart
  - stacked bar chart
  - mosaic plot
- determine which graphic(s) are appropriate to tell the story of your data
- interpret the presented graphics of others
- modify graphics to make them presentation-worthy

Code Preparation

To perform the code given in this slidedeck, please start R and run the following lines in the Script window in R:

```
source("http://rfs.kvasaheim.com/stat200.R")

dt = read.csv("http://rfs.kvasaheim.com/data/crime.csv")
attach(dt)
```


Three Graphics Theories

As an aside:

In R, there are three primary graphics engines (paradigms) that can be used. They use different ways of thinking about how the graphic should be created. The three engines are

- base
- grid
- ggplot

Our Graphics Metaphor

The one we will use in this class is `base`. The metaphor it uses seems much more natural to me:

The Painter's Canvas

1. plan your painting
2. define the parameters of the canvas
3. start the canvas
4. add to the canvas

For now, let us focus on creating basic graphics. Later today, we will see how to make them more presentable.

The Basic Categorical Graphics

Basic **uni**variate categorical graphics consist of

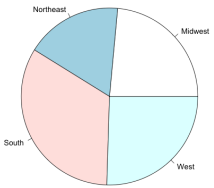
- Pie charts — categorical
- Bar charts — categorical

Basic **bi**variate categorical graphics consist of

- Side-by-side bar charts — categorical × categorical
- Stacked bar charts — categorical × categorical
- Spine charts — categorical × categorical
- Mosaic plots — categorical × categorical

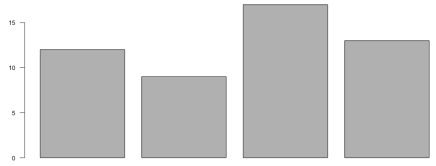
Pie Charts

```
pie( table(census4) )
```



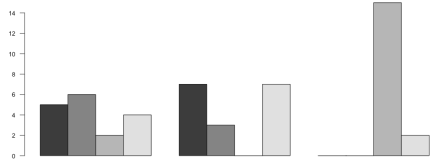
Bar Charts

```
barplot( table(census4) )
```



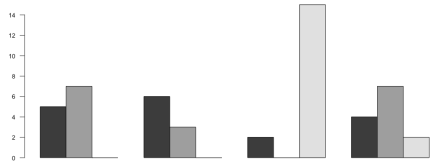
Side-by Side Bar Charts

```
barplot( table(census4,domPolCulture), beside=TRUE )
```



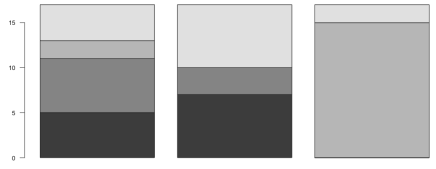
Side-by Side Bar Charts

```
barplot( table(domPolCulture,census4), beside=TRUE )
```



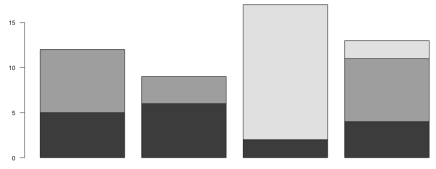
Stacked Bar Charts

```
barplot( table(census4, domPolCulture) )
```



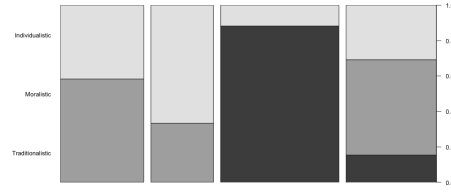
Stacked Bar Charts

```
barplot( table(domPolCulture, census4) )
```



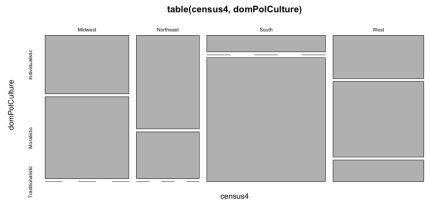
Stacked Bar Charts (a.k.a. Spine Chart)

```
spineplot(table(census4, domPolCulture))
```



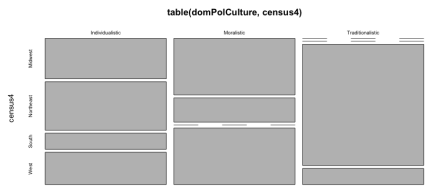
Mosaic Charts

```
mosaicplot( table(census4,domPolCulture) )
```



Mosaic Charts

```
mosaicplot( table(domPolCulture, census4) )
```



Modifying Graphics

Recall the metaphor for base graphics: **The Painter's Canvas**.

1. plan your painting
2. define the parameters of the canvas
3. start the canvas
4. add to the canvas

Note: Remember that the *only* purpose of graphics is to tell the "story of the data."

The Planning

Here are some things to think about before starting your graphic:

- what data should be illustrated by the graphic?
- where will the data be placed on the canvas?
- what margins will you use?
- what font? should it differ between labels and values?
- what sizes should the fonts be? should they differ between labels and values?
- should the axes have additional space?
- what should the value orientation be?

Note: It is very helpful to search the Internet for ideas for your graphic. I do this before I start the graphics for every major project.

The Parameters

Now that you have thought about what your graphic will look like (the planning), let's formally set the parameters of the graphic. The code at the bottom does the following:

- sets the margins to be 4 lines at the bottom, 3 at the left, 1 at the top, and 1 at the right;
- specifies the font family is serif, with labels bolded and values italicized;
- specifies the labels are 20% larger and values are 20% smaller than the base size;
- requires the axes to meet at $(0, 0)$; and
- forces all labels to be horizontal (for readability).

```
par( mar=c(4,3,1,1) )
par( family="serif", font.lab=2, font.axis=3 )
par( cex.lab=1.2, cex.axis=0.8 )
par( xaxs="i", yaxs="i" )
par( las=1 )
```

The Start

The following code:

- starts the graphic; and
- sets the viewing window to $(0, 10) \times (0, 100)$.

```
plot.new()
plot.window( xlim=c(0,10), ylim=c(0,100) )
```

This will start a generic window. Some graphing functions have their own "starting" functions. Those we are covering in this section do. So, the above two lines are not needed for those functions.

To create a bar chart with some interesting colors, we would replace these two lines with this one:

```
barplot( table(census4), col=c("red4", 3, "#291999", rgb(0.5,0.5,0.5)) )
```


The Annotation

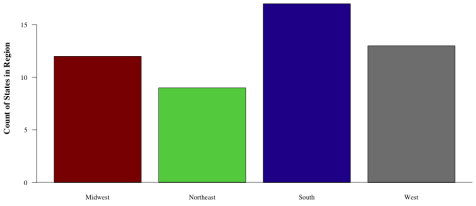
The following code:

- cleans up the graphic by adding label for the variables
- and adds a base line

```
title(xlab="Census Region", line=2.75)
title(ylab="Count of States in Region", line=2.75)
abline(h=0, lwd=2)
```

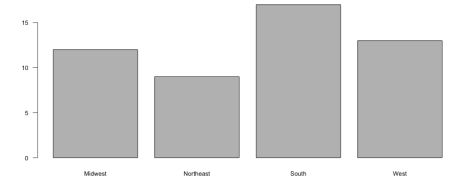
The Resulting Bar Chart

These all come together to produce this graphic:



The Resulting Bar Chart

Recall the original (utilitarian) graphic:



Today's Objectives

Now that we have concluded this lecture, you should be able to

- create the following univariate graphics for categorical variables:
  - pie chart
  - bar chart
- create the following bivariate graphics for categorical variables:
  - side-by-side bar chart
  - stacked bar chart
  - mosaic plot
- determine which graphic(s) are appropriate to tell the story of your data
- interpret the presented graphics of others
- modify graphics to make them presentation-worthy

Today's R Functions

In this slide deck, we covered the following R functions:

- `pie`
- `barplot`
- `spineplot`
- `mosaicplot`
- `par`
- `plot.new`
- `plot.window`
- `table`
- `abline`
- `title`

Supplemental Activities

The following may be of interest to you in terms of today's topics:

- SCA 3a is for **categorical** graphics
- SCA 3b is for numeric graphics

Note that you can access all Statistical Computing Activities here:
`https://www.kvasaheim.com/courses/stat200/sca/`

Supplemental Readings

The following are some readings that may be of interest to you in terms of graphing in R:

- R Graphics Cookbook, 2nd edition
  `https://r-graphics.org`
- R Coder
  `https://r-coder.com/r-graphs/`
- Statistical Methods and Data Analytics UCLA
  `https://stats.oarc.ucla.edu/r/codefragments/introduction/`
- Some colors and ideas:
  - `https://colorbrewer2.org/`
  - `https://r-charts.com/colors/`

Supplemental Readings

The following may be of interest to you in terms of today's topics:

- Hawkes Learning: Section 2.2
- Intro to Modern Statistics: Chapter 4
- R for Starters: Nothing