Chapter 4 Notes

Table of contents

# Chapter 4: Exploring US Census data with visualization

* **ggplot2** by Hadley Wickham is the core visualization framework for **tidyverse**
* Chapter goes over the following chart types:
  + faceted (“small multiples”)
  + population pyramids
  + Margin of Error plots
* Chapter goes over **plotly** for interactive vizualization

```{r setup}  
library(extrafont)  
```

Registering fonts with R

```{r setup}  
loadfonts(device = "win")  
```

Agency FB already registered with windowsFonts().

Aladin already registered with windowsFonts().

Algerian already registered with windowsFonts().

AquilineTwo already registered with windowsFonts().

Arial Black already registered with windowsFonts().

Arial already registered with windowsFonts().

Arial Narrow already registered with windowsFonts().

Arial Rounded MT Bold already registered with windowsFonts().

Avalon Quest already registered with windowsFonts().

Bahnschrift already registered with windowsFonts().

Baskerville Old Face already registered with windowsFonts().

Bauhaus 93 already registered with windowsFonts().

Bell MT already registered with windowsFonts().

Berlin Sans FB already registered with windowsFonts().

Berlin Sans FB Demi already registered with windowsFonts().

Bernard MT Condensed already registered with windowsFonts().

Blackadder ITC already registered with windowsFonts().

Blackout already registered with windowsFonts().

Blackout 2AM already registered with windowsFonts().

Blackout Midnight already registered with windowsFonts().

Blackout Sunrise already registered with windowsFonts().

Bodoni MT already registered with windowsFonts().

Bodoni MT Black already registered with windowsFonts().

Bodoni MT Condensed already registered with windowsFonts().

Bodoni MT Poster Compressed already registered with windowsFonts().

Book Antiqua already registered with windowsFonts().

Bookman Old Style already registered with windowsFonts().

Bookshelf Symbol 7 already registered with windowsFonts().

Bradley Hand ITC already registered with windowsFonts().

Bridgnorth already registered with windowsFonts().

Britannic Bold already registered with windowsFonts().

Broadway already registered with windowsFonts().

Brush Script MT already registered with windowsFonts().

Calibri already registered with windowsFonts().

Calibri Light already registered with windowsFonts().

Californian FB already registered with windowsFonts().

Calisto MT already registered with windowsFonts().

Cambria already registered with windowsFonts().

Candara already registered with windowsFonts().

Candara Light already registered with windowsFonts().

Cascadia Code already registered with windowsFonts().

Cascadia Mono already registered with windowsFonts().

Castellar already registered with windowsFonts().

Centaur already registered with windowsFonts().

Century already registered with windowsFonts().

Century Gothic already registered with windowsFonts().

Century Schoolbook already registered with windowsFonts().

Chiller already registered with windowsFonts().

ChunkFive already registered with windowsFonts().

Colonna MT already registered with windowsFonts().

Comic Sans MS already registered with windowsFonts().

Consolas already registered with windowsFonts().

Constantia already registered with windowsFonts().

Cooper Black already registered with windowsFonts().

Copperplate Gothic Bold already registered with windowsFonts().

Copperplate Gothic Light already registered with windowsFonts().

Corbel already registered with windowsFonts().

Corbel Light already registered with windowsFonts().

Courier New already registered with windowsFonts().

Curlz MT already registered with windowsFonts().

Dubai already registered with windowsFonts().

Dubai Light already registered with windowsFonts().

Dubai Medium already registered with windowsFonts().

Ebrima already registered with windowsFonts().

Edwardian Script ITC already registered with windowsFonts().

Elephant already registered with windowsFonts().

Engravers MT already registered with windowsFonts().

Eras Bold ITC already registered with windowsFonts().

Eras Demi ITC already registered with windowsFonts().

Eras Light ITC already registered with windowsFonts().

Eras Medium ITC already registered with windowsFonts().

Fanwood already registered with windowsFonts().

Fanwood Text already registered with windowsFonts().

Far East already registered with windowsFonts().

Felix Titling already registered with windowsFonts().

Footlight MT Light already registered with windowsFonts().

Forte already registered with windowsFonts().

Franklin Gothic Book already registered with windowsFonts().

Franklin Gothic Demi already registered with windowsFonts().

Franklin Gothic Demi Cond already registered with windowsFonts().

Franklin Gothic Heavy already registered with windowsFonts().

Franklin Gothic Medium already registered with windowsFonts().

Franklin Gothic Medium Cond already registered with windowsFonts().

Freestyle Script already registered with windowsFonts().

French Script MT already registered with windowsFonts().

Gabriola already registered with windowsFonts().

Gadugi already registered with windowsFonts().

Gaeilge 1 already registered with windowsFonts().

Garamond already registered with windowsFonts().

Georgia already registered with windowsFonts().

Gigi already registered with windowsFonts().

Gill Sans Ultra Bold already registered with windowsFonts().

Gill Sans Ultra Bold Condensed already registered with windowsFonts().

Gill Sans MT already registered with windowsFonts().

Gill Sans MT Condensed already registered with windowsFonts().

Gill Sans MT Ext Condensed Bold already registered with windowsFonts().

Gloucester MT Extra Condensed already registered with windowsFonts().

Goudy Bookletter 1911 already registered with windowsFonts().

GoudyMedieval already registered with windowsFonts().

Goudy Old Style already registered with windowsFonts().

Sorts Mill Goudy already registered with windowsFonts().

Goudy Stout already registered with windowsFonts().

Haettenschweiler already registered with windowsFonts().

Harlow Solid Italic already registered with windowsFonts().

Harrington already registered with windowsFonts().

High Tower Text already registered with windowsFonts().

HoloLens MDL2 Assets already registered with windowsFonts().

IM FELL English already registered with windowsFonts().

Impact already registered with windowsFonts().

Imprint MT Shadow already registered with windowsFonts().

Informal Roman already registered with windowsFonts().

Ink Free already registered with windowsFonts().

Javanese Text already registered with windowsFonts().

Jokerman already registered with windowsFonts().

Juice ITC already registered with windowsFonts().

Junction Bold already registered with windowsFonts().

Junction Light already registered with windowsFonts().

Junction Regular already registered with windowsFonts().

Knewave already registered with windowsFonts().

Knewave Outline already registered with windowsFonts().

Kristen ITC already registered with windowsFonts().

Kunstler Script already registered with windowsFonts().

Wide Latin already registered with windowsFonts().

League Script Thin already registered with windowsFonts().

League Spartan Black already registered with windowsFonts().

League Spartan already registered with windowsFonts().

League Spartan ExtraBold already registered with windowsFonts().

League Spartan ExtraLight already registered with windowsFonts().

League Spartan Light already registered with windowsFonts().

League Spartan Medium already registered with windowsFonts().

League Spartan SemiBold already registered with windowsFonts().

Leelawadee already registered with windowsFonts().

Leelawadee UI already registered with windowsFonts().

Leelawadee UI Semilight already registered with windowsFonts().

Linden Hill already registered with windowsFonts().

Lucida Bright already registered with windowsFonts().

Lucida Calligraphy already registered with windowsFonts().

Lucida Console already registered with windowsFonts().

Lucida Fax already registered with windowsFonts().

Lucida Handwriting already registered with windowsFonts().

Lucida Sans already registered with windowsFonts().

Lucida Sans Typewriter already registered with windowsFonts().

Lucida Sans Unicode already registered with windowsFonts().

Magneto already registered with windowsFonts().

Maiandra GD already registered with windowsFonts().

Malgun Gothic already registered with windowsFonts().

Malgun Gothic Semilight already registered with windowsFonts().

Marlett already registered with windowsFonts().

MasonSerifAltBold already registered with windowsFonts().

MasonSerifAltReg already registered with windowsFonts().

MasonSerifBold already registered with windowsFonts().

MasonSerifRegular already registered with windowsFonts().

MasonSerifSuper already registered with windowsFonts().

MasonSerifSuperBold already registered with windowsFonts().

Matura MT Script Capitals already registered with windowsFonts().

Microsoft Himalaya already registered with windowsFonts().

Microsoft Yi Baiti already registered with windowsFonts().

Microsoft New Tai Lue already registered with windowsFonts().

Microsoft PhagsPa already registered with windowsFonts().

Microsoft Sans Serif already registered with windowsFonts().

Microsoft Tai Le already registered with windowsFonts().

Microsoft Uighur already registered with windowsFonts().

Middle Ages already registered with windowsFonts().

Mistral already registered with windowsFonts().

Modern No. 20 already registered with windowsFonts().

Mongolian Baiti already registered with windowsFonts().

Monotype Corsiva already registered with windowsFonts().

MS Outlook already registered with windowsFonts().

MS Reference Sans Serif already registered with windowsFonts().

MS Reference Specialty already registered with windowsFonts().

MT Extra already registered with windowsFonts().

MV Boli already registered with windowsFonts().

Myanmar Text already registered with windowsFonts().

Niagara Engraved already registered with windowsFonts().

Niagara Solid already registered with windowsFonts().

Nirmala UI already registered with windowsFonts().

Nirmala UI Semilight already registered with windowsFonts().

OCR A Extended already registered with windowsFonts().

OFL Sorts Mill Goudy TT already registered with windowsFonts().

Old English Text MT already registered with windowsFonts().

Onyx already registered with windowsFonts().

Orbitron already registered with windowsFonts().

Ostrich Sans already registered with windowsFonts().

Ostrich Sans Inline already registered with windowsFonts().

Ostrich Sans Rounded already registered with windowsFonts().

Palace Script MT already registered with windowsFonts().

Palatino Linotype already registered with windowsFonts().

Papyrus already registered with windowsFonts().

Parchment already registered with windowsFonts().

Perpetua already registered with windowsFonts().

Perpetua Titling MT already registered with windowsFonts().

Playbill already registered with windowsFonts().

Poor Richard already registered with windowsFonts().

Pristina already registered with windowsFonts().

Prociono already registered with windowsFonts().

Prociono TT already registered with windowsFonts().

Rage Italic already registered with windowsFonts().

Raleway Black already registered with windowsFonts().

Raleway already registered with windowsFonts().

Raleway ExtraBold already registered with windowsFonts().

Raleway ExtraLight already registered with windowsFonts().

Raleway Light already registered with windowsFonts().

Raleway Medium already registered with windowsFonts().

Raleway SemiBold already registered with windowsFonts().

Raleway Thin already registered with windowsFonts().

Ravie already registered with windowsFonts().

Rockwell already registered with windowsFonts().

Rockwell Condensed already registered with windowsFonts().

Rockwell Extra Bold already registered with windowsFonts().

Sans Serif Collection already registered with windowsFonts().

Script MT Bold already registered with windowsFonts().

Segoe Fluent Icons already registered with windowsFonts().

Segoe MDL2 Assets already registered with windowsFonts().

Segoe Print already registered with windowsFonts().

Segoe Script already registered with windowsFonts().

Segoe UI already registered with windowsFonts().

Segoe UI Light already registered with windowsFonts().

Segoe UI Semibold already registered with windowsFonts().

Segoe UI Semilight already registered with windowsFonts().

Segoe UI Black already registered with windowsFonts().

Segoe UI Historic already registered with windowsFonts().

Segoe UI Symbol already registered with windowsFonts().

Segoe UI Variable already registered with windowsFonts().

Shanghai already registered with windowsFonts().

Showcard Gothic already registered with windowsFonts().

SimSun-ExtB already registered with windowsFonts().

Sitka Text already registered with windowsFonts().

Snap ITC already registered with windowsFonts().

Sniglet already registered with windowsFonts().

Stencil already registered with windowsFonts().

Sylfaen already registered with windowsFonts().

Symbol already registered with windowsFonts().

Tahoma already registered with windowsFonts().

Tempus Sans ITC already registered with windowsFonts().

Times New Roman already registered with windowsFonts().

Tolkien already registered with windowsFonts().

Trebuchet MS already registered with windowsFonts().

Tw Cen MT already registered with windowsFonts().

Tw Cen MT Condensed already registered with windowsFonts().

Tw Cen MT Condensed Extra Bold already registered with windowsFonts().

Verdana already registered with windowsFonts().

Viner Hand ITC already registered with windowsFonts().

Vivaldi already registered with windowsFonts().

Vladimir Script already registered with windowsFonts().

Webdings already registered with windowsFonts().

Wingdings already registered with windowsFonts().

Wingdings 2 already registered with windowsFonts().

Wingdings 3 already registered with windowsFonts().

Worn Manuscript already registered with windowsFonts().

```{r setup}  
library(tidycensus)  
library(tidyverse)  
```

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.2 ✔ readr 2.1.4  
✔ forcats 1.0.0 ✔ stringr 1.5.0  
✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
✔ purrr 1.0.2

── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

## 4.1 Basic Census visualization with ggplot2

* visualization is the examination of patterns and trends in data graphically
* visualization is especially key for exploratory analysis
* Below we’ll pull in the data we’ll work through - 2020 5-year Georgia median household income and median age by county, in a wide table format

```{r mhi-georgia}  
ga\_wide <- get\_acs(  
 geography = "county",  
 state = "Georgia",  
 variables = c(medinc = "B19013\_001",  
 medage = "B01002\_001"),  
 output = "wide",  
 year = 2020  
)  
```

Getting data from the 2016-2020 5-year ACS

```{r mhi-georgia}  
ga\_wide  
```

# A tibble: 159 × 6  
 GEOID NAME medincE medincM medageE medageM  
 <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
 1 13001 Appling County, Georgia 37924 4761 39.9 1.7  
 2 13003 Atkinson County, Georgia 35703 5493 35.9 1.5  
 3 13005 Bacon County, Georgia 36692 3774 36.5 1   
 4 13007 Baker County, Georgia 34034 9879 52.2 4.8  
 5 13011 Banks County, Georgia 50912 4278 41.5 1.1  
 6 13013 Barrow County, Georgia 62990 2562 36 0.3  
 7 13017 Ben Hill County, Georgia 32077 4008 39.5 1.4  
 8 13021 Bibb County, Georgia 41317 1220 36.3 0.3  
 9 13023 Bleckley County, Georgia 46992 6279 36 1.5  
10 13027 Brooks County, Georgia 37516 4438 43.6 0.9  
# ℹ 149 more rows

### Getting started with ggplot2

1. Initalize **ggplot2** visualizations as a plot object using ggplot()
   * First argument is a dataset to be visualized
   * Second argument is default aes() *(aesthetic)* mappings for the plot
     + mappings can be applied to elements such as data axes, fill, color, etc.

After initializing the plot object, we layer plot elements onto the object

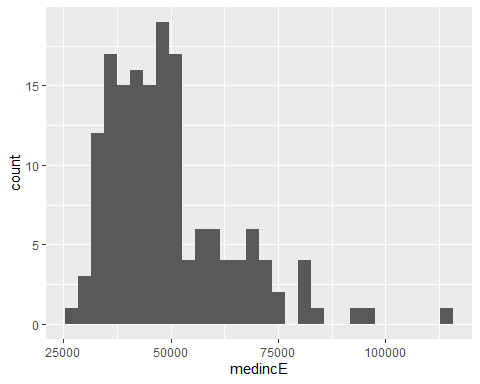
* Main factor here is the geom, determining the chart type to be used.
  + geom\_bar() - bar plots
  + geom\_line() - line plots
  + geom\_point() - point plots
  + geom\_histogram() - histograms
  + geom\_boxplot() - box and whisker plots
* To add a geom layer, use the + operator

#### Histograms

We’ll start with a histogram to look at data distribution

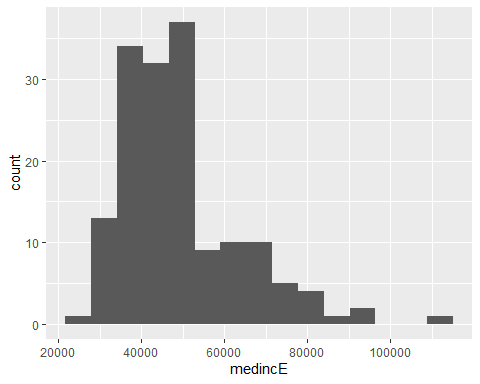
```{r georgia-histo}  
options(scipen = 999) # optional call; avoid scientific notation  
  
ggplot(ga\_wide, aes(x = medincE)) +  
 geom\_histogram()  
```

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



* A histogram displays the distribution of data values
  + the horizontal axis represents the range of values from least to greatest
  + the vertical axis represents the count, or number quantity of data points with a given value
    - a given range of values is ‘binned’ into a box whose height represent the count of values that may be classed in said box
    - by default, **ggplot2** organizes the data values into 30 bins
      * this can be changed via the bins parameter in geom\_histogram(), demonstrated below

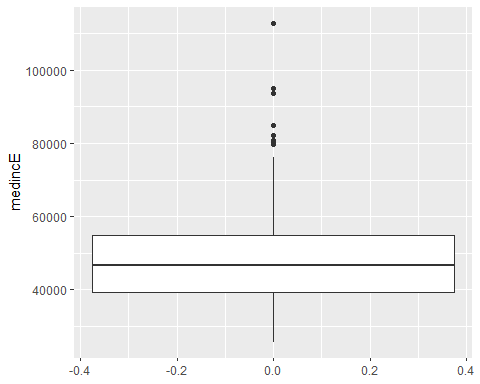
```{r georgia-15bin-histo}  
ggplot(ga\_wide, aes(x = medincE)) +  
 geom\_histogram(bins = 15)  
```



#### Box-and-Whisker Plot

* The box-and-whisker plot is a common alternative to the histogram
  + The central box of a box plot shows the *interquartile range (IQR)* or those values between the 25th and 75th percentile
    - The central line of the box represents the distribution median
  + the *whiskers* demonstrate values outside the IQR as follows:
    - they may stretch to the min/max values of the distribution, OR
    - they may represent 1.5x the IQR in either, or both, directions
  + values plotted as points plotted beyond the whiskers are *outliers*

```{r georgia-boxplot}  
ggplot(ga\_wide, aes(y = medincE)) +  
 geom\_boxplot()  
```

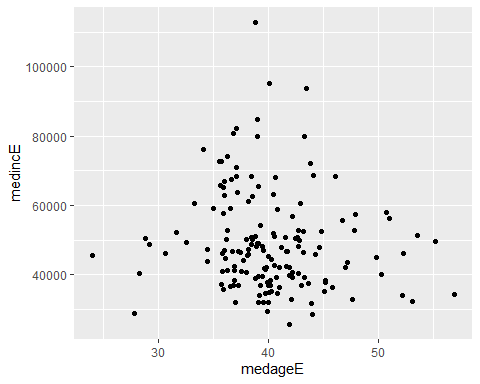


* in the above boxplot:
  + the aes() call maps the plot to the y axis, creating a vertical boxplot
  + the lower whiskers stretches to the minimum value
  + the upper whiskers stretches to 1.5x the IQR
  + we see several points plotted as outliers above the upper bound

#### (Scatter plots) Visualizing multivariate relationships

* We may want to explore/visualize interrelationship between two (or more) variables
* with two numeric variables, we may explore relationships using *scatter plots*
  + one column of values is mapped to the X-axis
  + the other is mapped to the Y-axis
  + geom\_point() makes scatter plots
    - two columns are needed for the aes() mapping

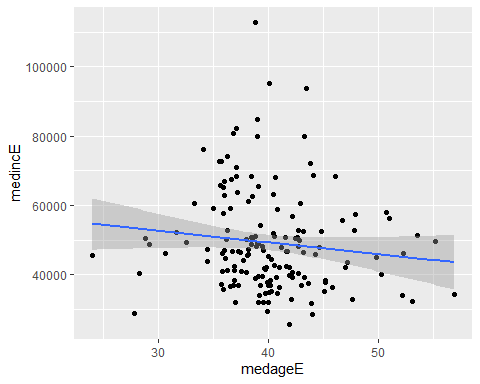
```{r georgia-medage-medinc-scatter}  
ggplot(ga\_wide, aes(x = medageE, y = medincE)) +  
 geom\_point()  
```



* Sometimes, a plot like above clearly demonstrates the existence of a relationship between variables (*correlation)*
* Sometimes, the plot fails to make such correlation immediately clear
  + in these cases, we may add further elements to clarify
  + geom\_smooth(), for instance, draws a fitted line representing the relationship between the plotted columns
    - method = "lm" uses the linear model to fit a line
    - method = "loess" uses Local Polynomial Regression Fitting to fit smoothed relationships

```{r georgia-medage-medinc-scatter-lm-smooth}  
ggplot(ga\_wide, aes(x = medageE, y = medincE)) +  
 geom\_point() +  
 geom\_smooth(method = "lm")  
```

`geom\_smooth()` using formula = 'y ~ x'



* the regression line has a light downward movement from left to right, implying somewhat of a negative relationship between the two variables
  + we might interpret this as median household income declining slightly as median age increases

## 4.2 Customizing ggplot2 visualizations

* **ggplot2** provides useful and attractive defaults, but analysts will want to customize visualizations for presenting to diverse audiences
* The subsection follows an example of visualization prep for data showing percent of commuters who use urban public transit
  + we take this data from 2019 1-year ACS Data Profile
  + The variable we use is DP03\_0021P - “Percent!!COMMUTING TO WORK!!Workers 16 years and over!!Public transportation (excluding taxicab)”
  + We’ll find the 20 largest metropolitan areas by pop using slice\_max()

```{r metro-pt-commute}  
metros <- get\_acs(  
 geography = "cbsa",  
 variables = "DP03\_0021P",  
 summary\_var = "B01003\_001",  
 survey = "acs1",  
 year = 2019  
) %>%   
 slice\_max(summary\_est, n = 20)  
```

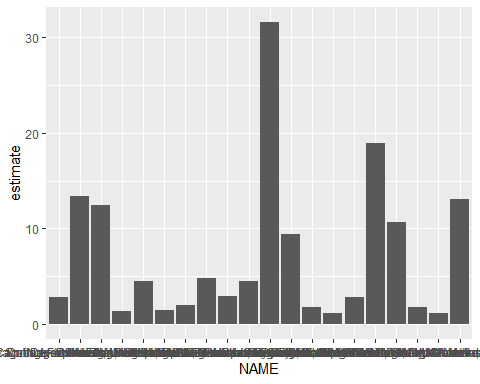
Getting data from the 2019 1-year ACS

The 1-year ACS provides data for geographies with populations of 65,000 and greater.

Using the ACS Data Profile

* data returned is 20x7, with two extra columns from the summary\_var and its associated summary\_moe, and the rows being the slice\_max return
* below, we create a bar plot comparing public transit commute share between metro areas

```{r initial-commute-chart}  
ggplot(metros, aes(x = NAME, y = estimate)) +  
 geom\_col()  
```



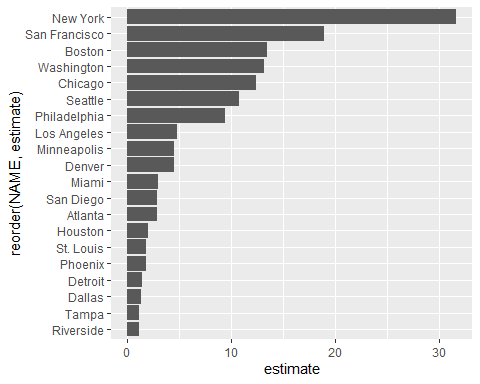
* the above plot is hard to interpret without fine-tuning the formatting
  + x-axis overlap, making reading impossible
  + axis titles are not intuitive
  + data isn’t sorted

### Improving legibility

* we can perform the above cleaning
  + **ggplot2** works with **magrittr** piping and **tidyverse** functions, meaning we can combine manipulation and visualization tasks

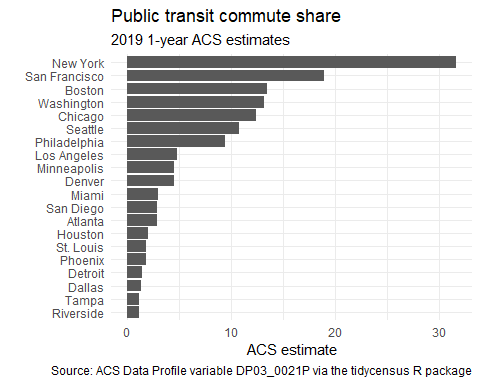
1. We will format NAME to be more intuitive using **stringr** info nested in mutate() calls - str\_remove() lets us use regex to remove unneeded parts of the NAME text for each area - the mutate calls are piped into ggplot() and ggplot() infers the dataset from the preceding pipe chains
2. We will improve legibility by putting metro name on y-axis and ACS estimate on x-axis
3. We will order data in descending order of estimated values using reorder() inside aes()
4. We plot it as a column plot

```{r cleaned-commute-chart-a}  
metros %>%   
 mutate(NAME = str\_remove(NAME, "-.\*$")) %>%   
 mutate(NAME = str\_remove(NAME, ",.\*$")) %>%   
 ggplot(aes(y = reorder(NAME, estimate), x = estimate)) +  
 geom\_col()  
```



* the above is much cleaner, but we need to fix the labels to make the chart interpretable
  + We can finish cleaning up by using labs() to specify labels
    - In this instance, we will include a title label, a subtitle label, we will leave the y axis label empty, and we will add a caption label containing data credits
  + We can also include themes (in this case, using theme\_minimal()

```{r final-commute-chart}  
metros %>%   
 mutate(NAME = str\_remove(NAME, "-.\*$")) %>%   
 mutate(NAME = str\_remove(NAME, ",.\*$")) %>%   
 ggplot(aes(y = reorder(NAME, estimate), x = estimate)) +  
 geom\_col() +  
 theme\_minimal() +  
 labs(title = "Public transit commute share",  
 subtitle = "2019 1-year ACS estimates",  
 y = "",  
 x = "ACS estimate",  
 caption = "Source: ACS Data Profile variable DP03\_0021P via the tidycensus R package")  
```



### Custom styling of ggplot2 charts

**ggplot2** maintains a wide scope for customization. Below, we demonstrate several style changes to color, transparency, font, and tick labels.

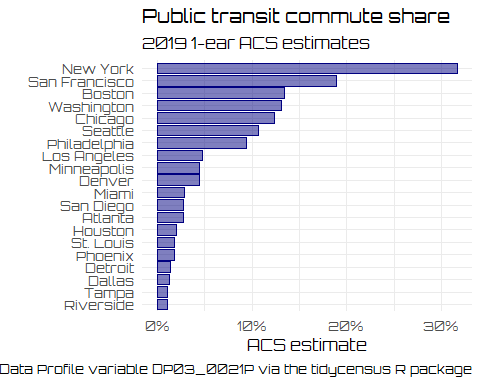
```{r customized-commute-chart}  
library(scales)  
```

Attaching package: 'scales'

The following object is masked from 'package:purrr':  
  
 discard

The following object is masked from 'package:readr':  
  
 col\_factor

```{r customized-commute-chart}  
metros %>%   
 mutate(NAME = str\_remove(NAME, "-.\*$")) %>%   
 mutate(NAME = str\_remove(NAME, ",.\*$")) %>%   
 ggplot(aes(y = reorder(NAME, estimate), x = estimate)) +  
 geom\_col(color = "navy", fill = "navy",  
 alpha = 0.5, width = 0.85) +  
 theme\_minimal(base\_size = 12, base\_family = "Orbitron") +  
 scale\_x\_continuous(labels = label\_percent(scale = 1)) +  
 labs(title = "Public transit commute share",  
 subtitle = "2019 1-ear ACS estimates",  
 y = "",  
 x = "ACS estimate",  
 caption = "Source: ACS Data Profile variable DP03\_0021P via the tidycensus R package")  
```



* The above code used the following modifications:
  + geom\_cols() had aes() mappings modified locally, rather than at the global level of the object initialization
    - the bar was given a color (the outline color) and fill (the fill color)
    - transparency was modified using alpha argument
    - width = 0.85 increased the space between bars by reducing the visual width of each bar to 85% of the default value
  + theme\_minimal() uses the base\_size and base\_family parameters, respectively
    - base\_size controls the font size of plot elements, defaulting to 11
    - base\_family designates the font family to use
      * in this case, the “Verdana” plot is used
      * any font family accessible from the operating system may be used
      * Check installed fonts with systemfonts::system\_fonts()