

# Lab 7

AUTHOR  
Conie O'Malley

PUBLISHED  
March 2, 2025

## 1 Lab Instructions

### 1.1 Deliverable 1: Get your working directory and paste below:

► Code

```
[1] "/Users/coniecakes/Library/CloudStorage/OneDrive-Personal/001. Documents - Main/023. Programming Tools/R Studio/AdvTxtAnalytics/Lab_7"
```

### 1.2 Deliverable 2: Load the Required libraries

► Code

```
Loading required package: geojsonio
```

```
Registered S3 method overwritten by 'geojsonsf':
```

```
  method      from  
  print.geojson geojson
```

```
Attaching package: 'geojsonio'
```

```
The following object is masked from 'package:base':
```

```
  pretty
```

```
Loading required package: ggmap
```

```
Loading required package: ggplot2
```

```
i Google's Terms of Service: <https://mapsplatform.google.com>
```

```
Stadia Maps' Terms of Service: <https://stadiamaps.com/terms-of-service/>
```

```
OpenStreetMap's Tile Usage Policy:
```

```
<https://operations.osmfoundation.org/policies/tiles/>
```

```
i Please cite ggmap if you use it! Use `citation("ggmap")` for details.
```

```
Loading required package: highcharter
```

```
Registered S3 method overwritten by 'quantmod':
```

```
  method      from  
  as.zoo.data.frame zoo
```

Loading required package: leaflet

Loading required package: plotly

Attaching package: 'plotly'

The following object is masked from 'package:ggmap':

wind

The following object is masked from 'package:ggplot2':

last\_plot

The following object is masked from 'package:stats':

filter

The following object is masked from 'package:graphics':

layout

Loading required package: Rserve

Warning: package 'Rserve' was built under R version 4.3.3

Loading required package: sf

Warning: package 'sf' was built under R version 4.3.3

Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf\_use\_s2() is TRUE

Loading required package: ggthemes

Loading required package: DT

Loading required package: gapminder

Loading required package: devtools

Loading required package: usethis

Warning: package 'usethis' was built under R version 4.3.3

Loading required package: remotes

Warning: package 'remotes' was built under R version 4.3.3

Attaching package: 'remotes'

The following objects are masked from 'package:devtools':

```
dev_package_deps, install_bioc, install_bitbucket, install_cran,  
install_deps, install_dev, install_git, install_github,  
install_gitlab, install_local, install_svn, install_url,  
install_version, update_packages
```

Loading required package: dotenv

Warning: package 'dotenv' was built under R version 4.3.3

## 2 Part 1: Data Visualization Using ggplot2

### 2.1 Deliverable 3: Get and Explore the Diamonds Dataset

► Code

```
# A tibble: 6 × 10  
  carat cut      color clarity depth table price     x     y     z  
  <dbl> <ord>    <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>  
1 0.23 Ideal    E      SI2     61.5    55   326  3.95  3.98  2.43  
2 0.21 Premium  E      SI1     59.8    61   326  3.89  3.84  2.31  
3 0.23 Good     E      VS1     56.9    65   327  4.05  4.07  2.31  
4 0.29 Premium  I      VS2     62.4    58   334  4.2   4.23  2.63  
5 0.31 Good     J      SI2     63.3    58   335  4.34  4.35  2.75  
6 0.24 Very Good J      VVS2    62.8    57   336  3.94  3.96  2.48
```

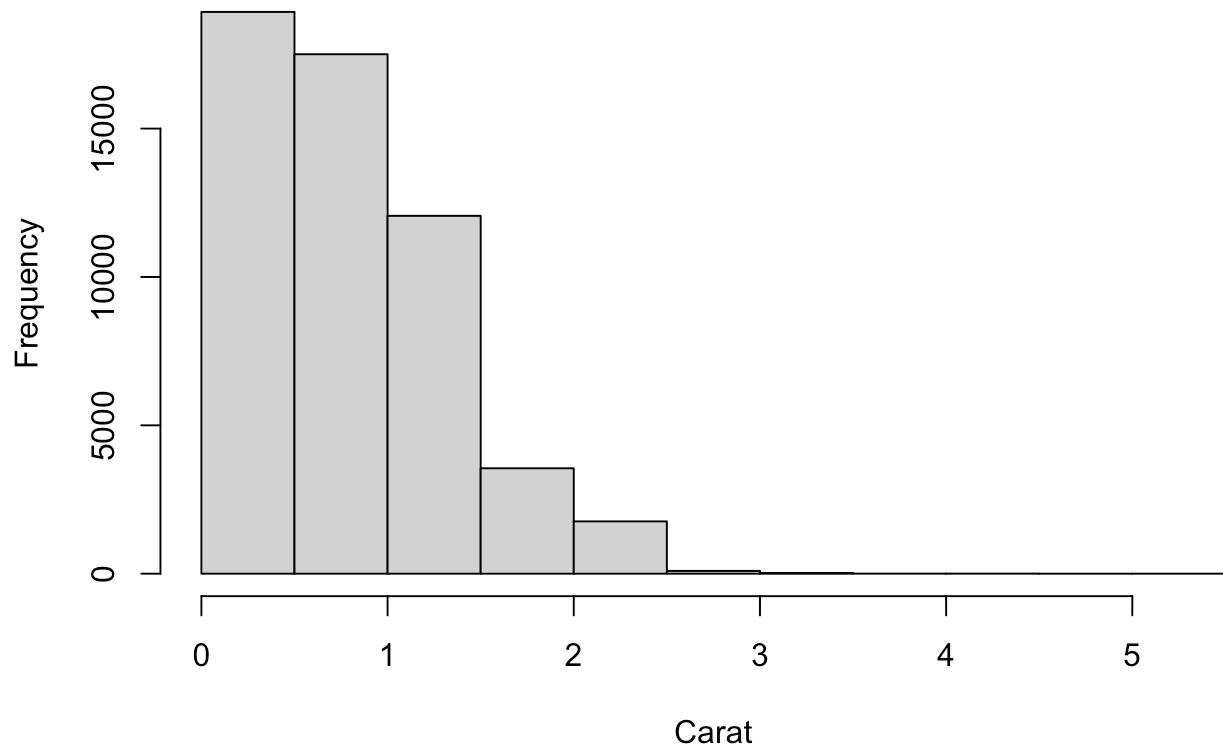
► Code

```
# A tibble: 6 × 10  
  carat cut      color clarity depth table price     x     y     z  
  <dbl> <ord>    <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>  
1 0.72 Premium  D      SI1     62.7    59   2757 5.69  5.73  3.58  
2 0.72 Ideal    D      SI1     60.8    57   2757 5.75  5.76  3.5  
3 0.72 Good     D      SI1     63.1    55   2757 5.69  5.75  3.61  
4 0.7  Very Good D      SI1     62.8    60   2757 5.66  5.68  3.56  
5 0.86 Premium  H      SI2     61      58   2757 6.15  6.12  3.74  
6 0.75 Ideal    D      SI2     62.2    55   2757 5.83  5.87  3.64
```

### 2.2 Deliverable 4: Create a Histogram of "diamonds" in Base R

► Code

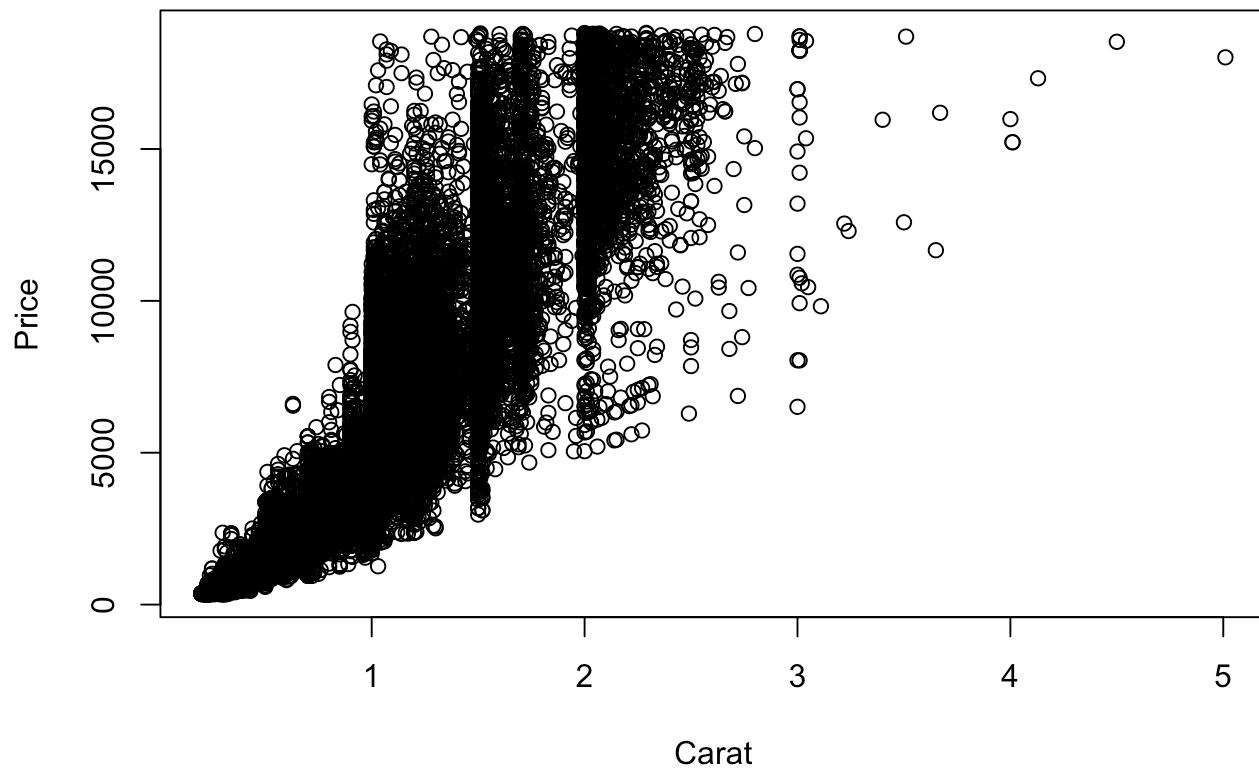
### Carat Histogram



## 2.3 Deliverable 5: Create a Scatterplot of "diamonds" using Base R

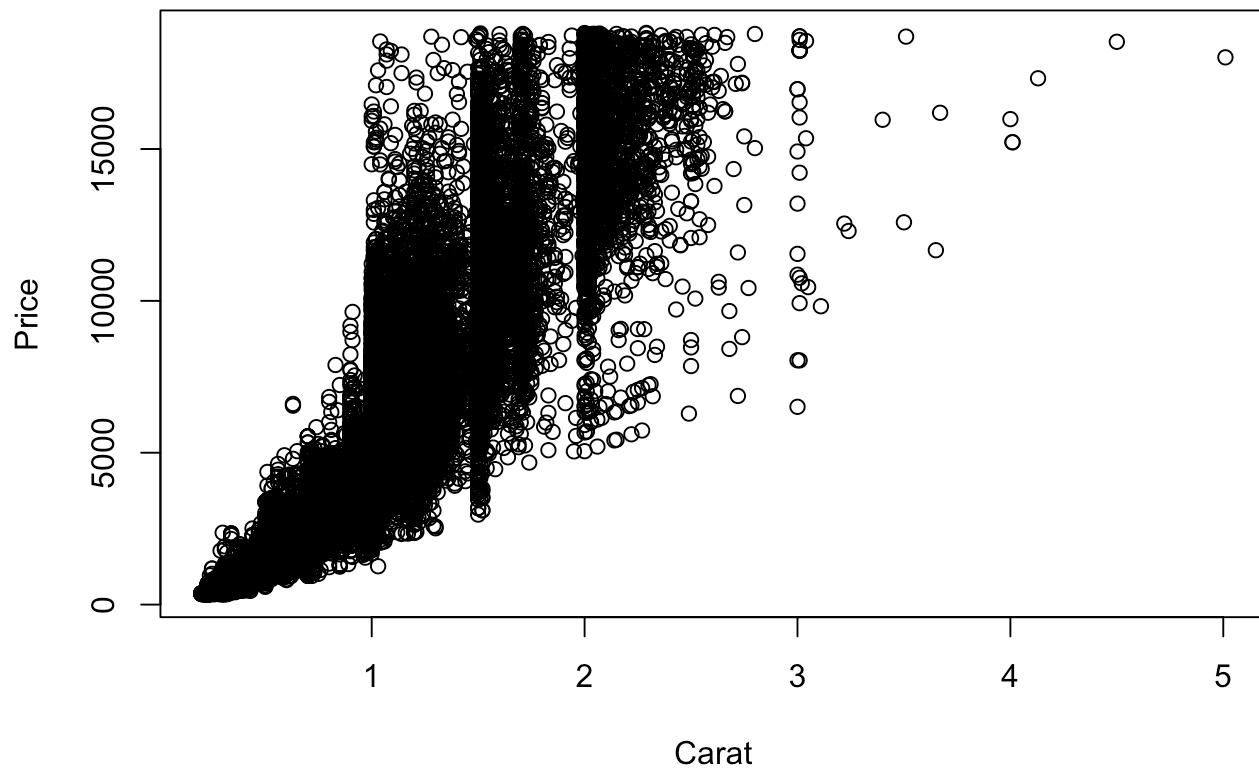
► Code

### Scatterplot of Carat vs Price



► [Code](#)

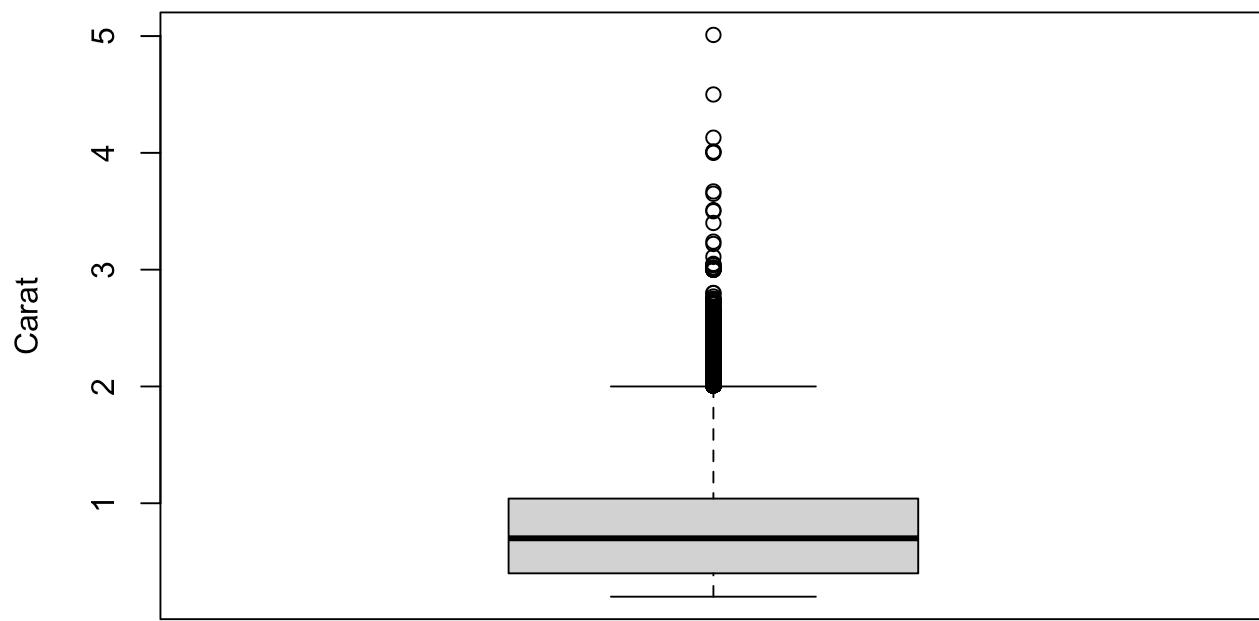
### Scatterplot of Carat vs Price



## 2.4 Deliverable 6: Build a Boxplot of carat in the diamonds dataset

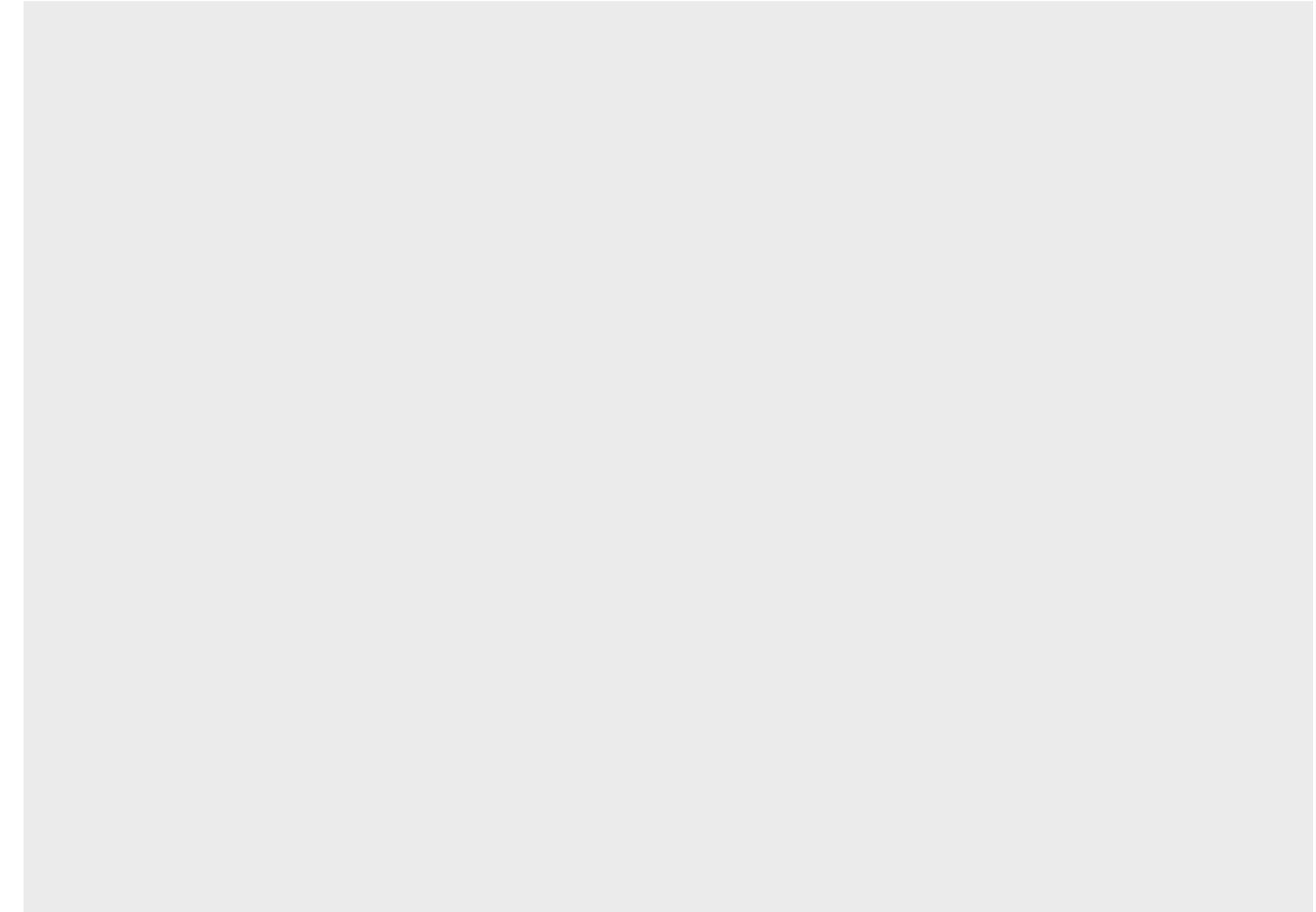
► [Code](#)

### Boxplot of Diamonds by Carat



## 2.5 Deliverable 7: Create a Blank Canvas in ggplot

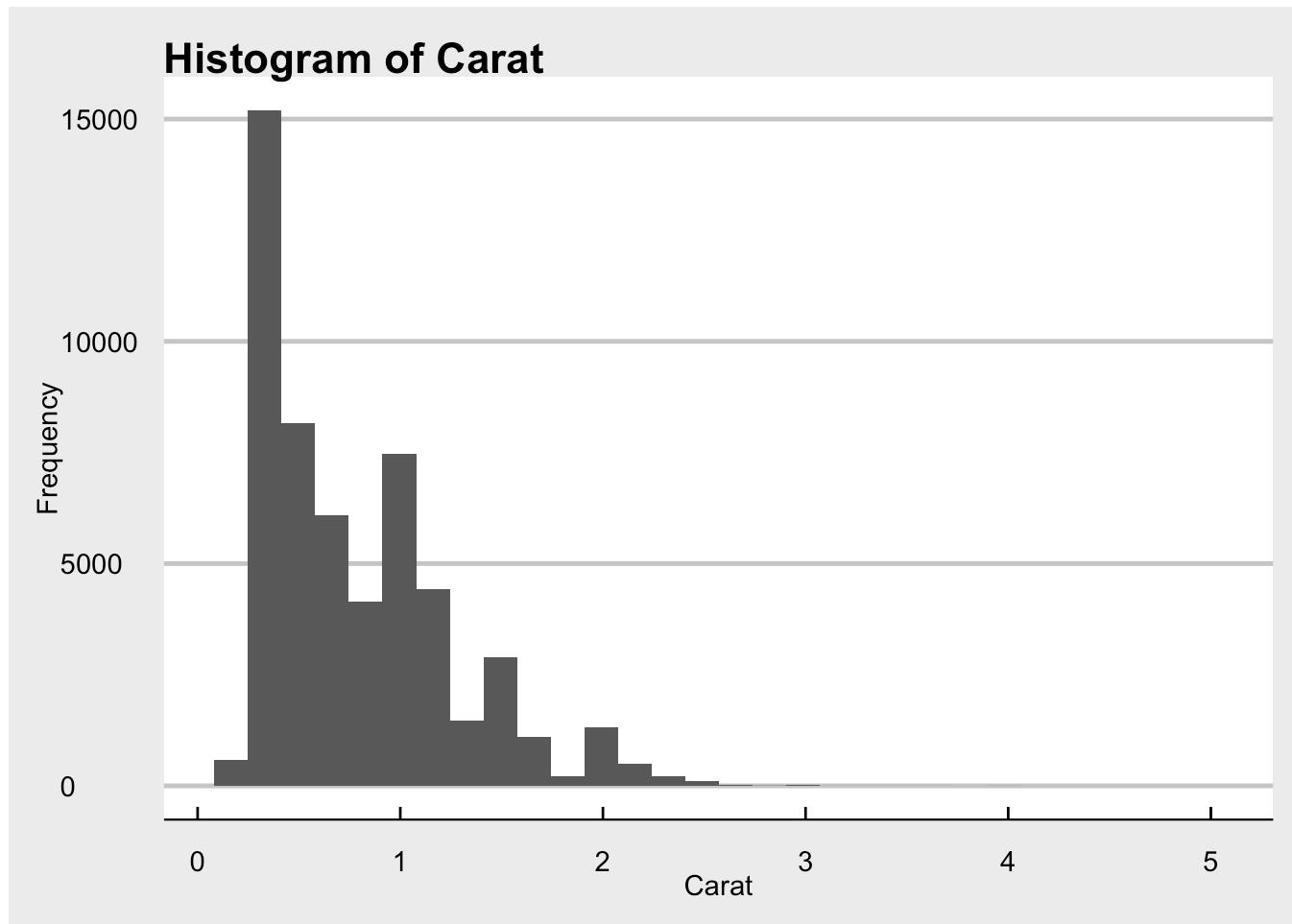
► Code



## 2.6 Deliverable 8: Rebuild the diamonds histogram with ggplot

► Code

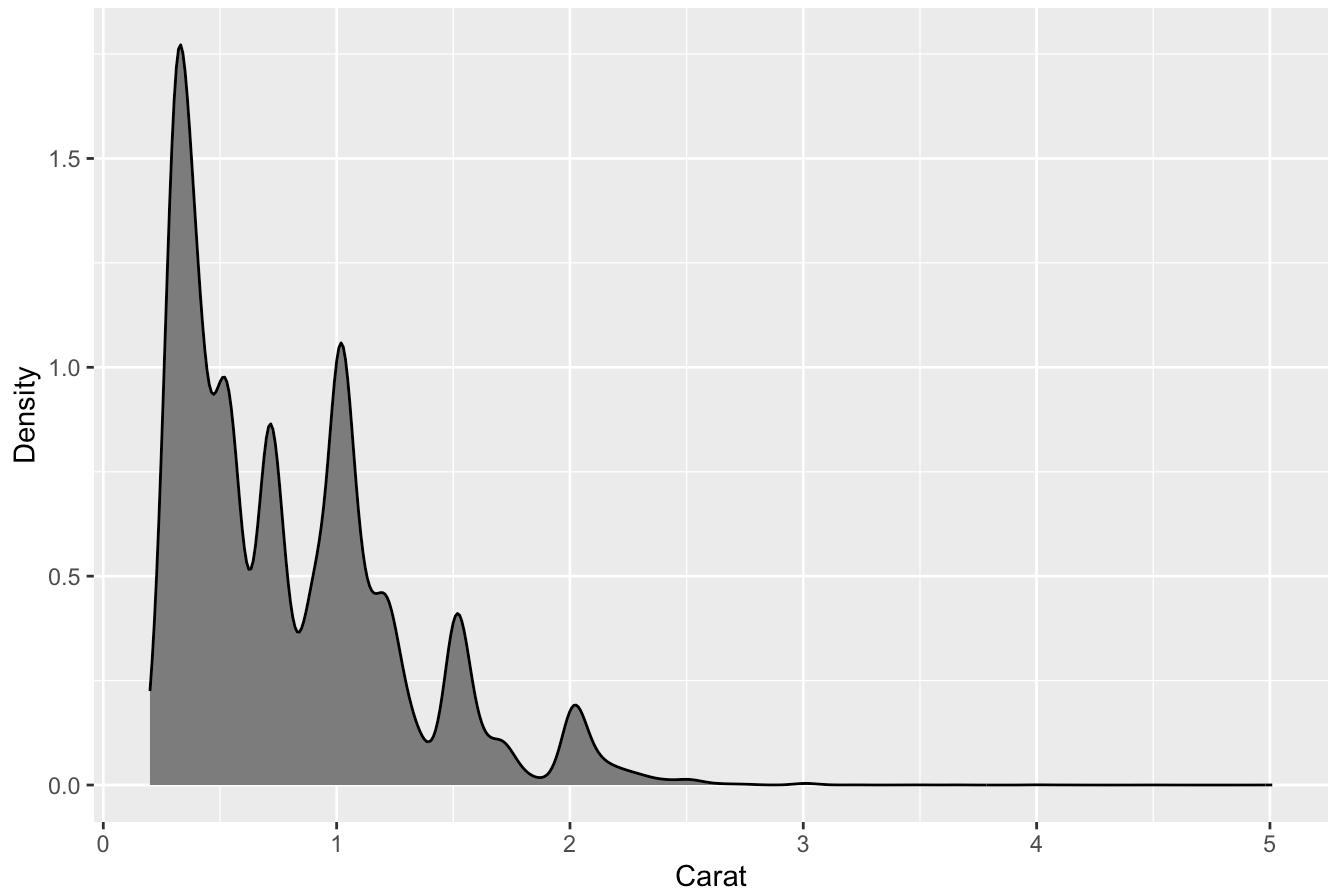
```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



## 2.7 Deliverable 9: Build a Density Plot

► Code

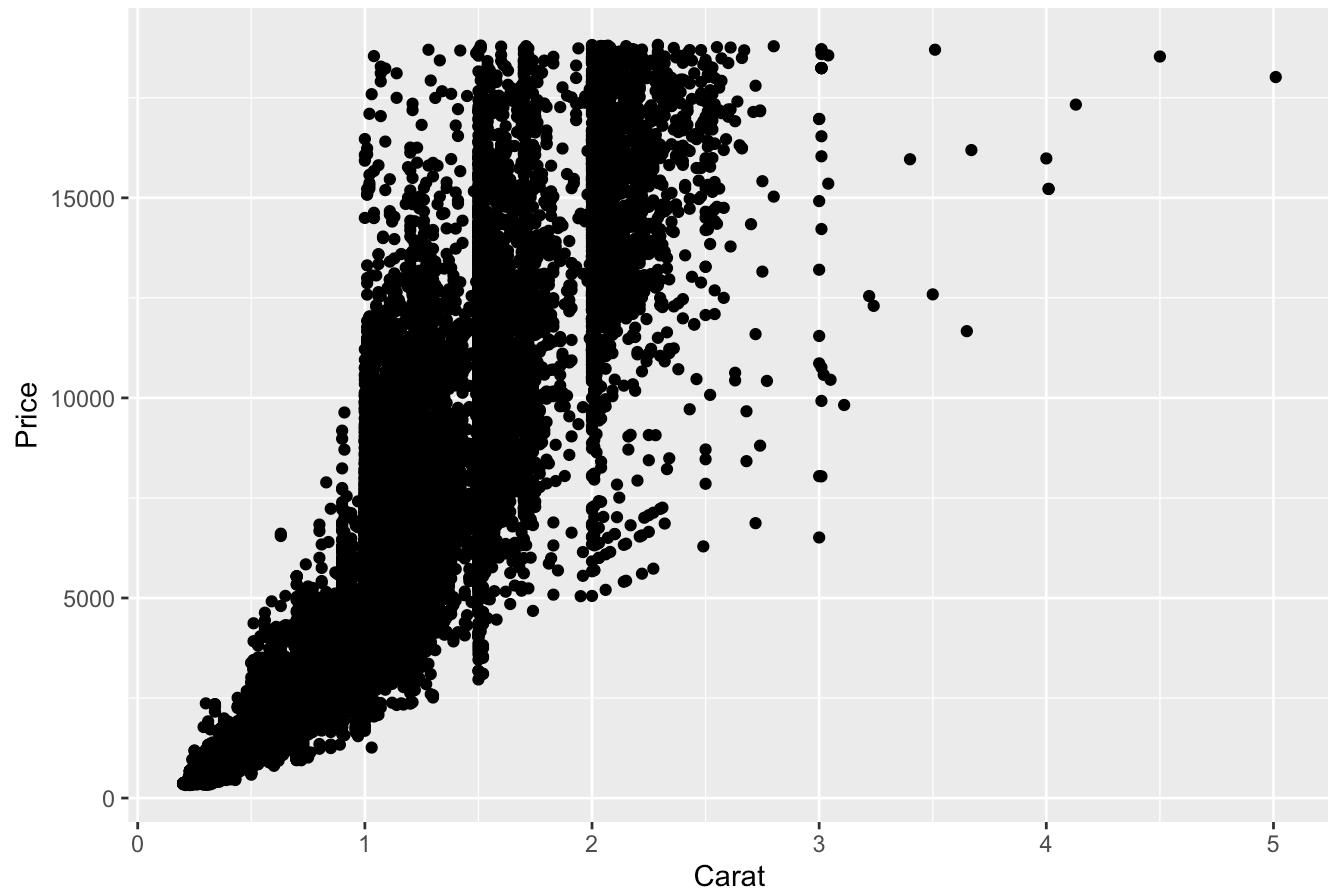
Density plot of Carat



## 2.8 Deliverable 10: Build a Scatterplot with ggplot

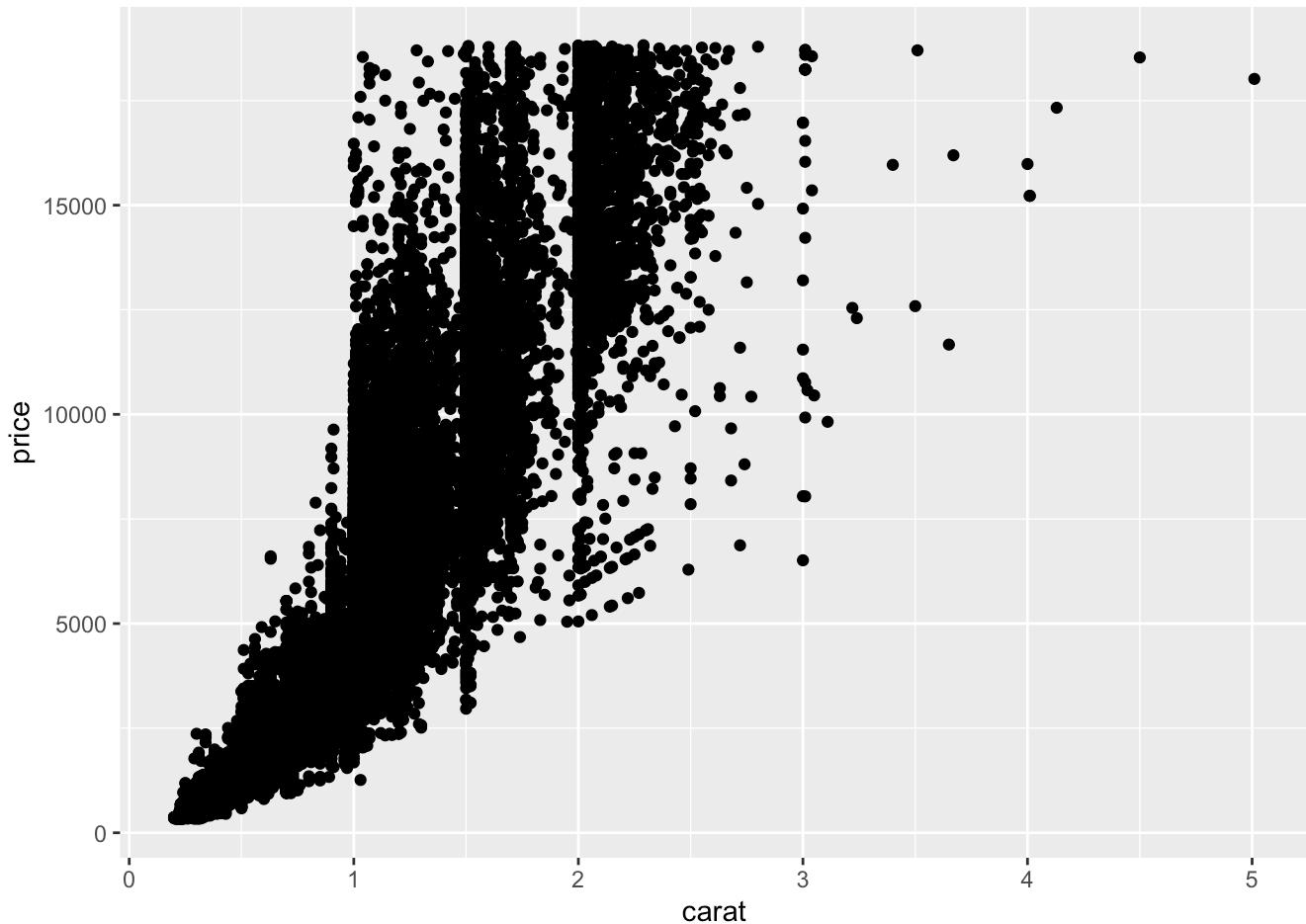
► Code

Scatterplot of Carat vs Price



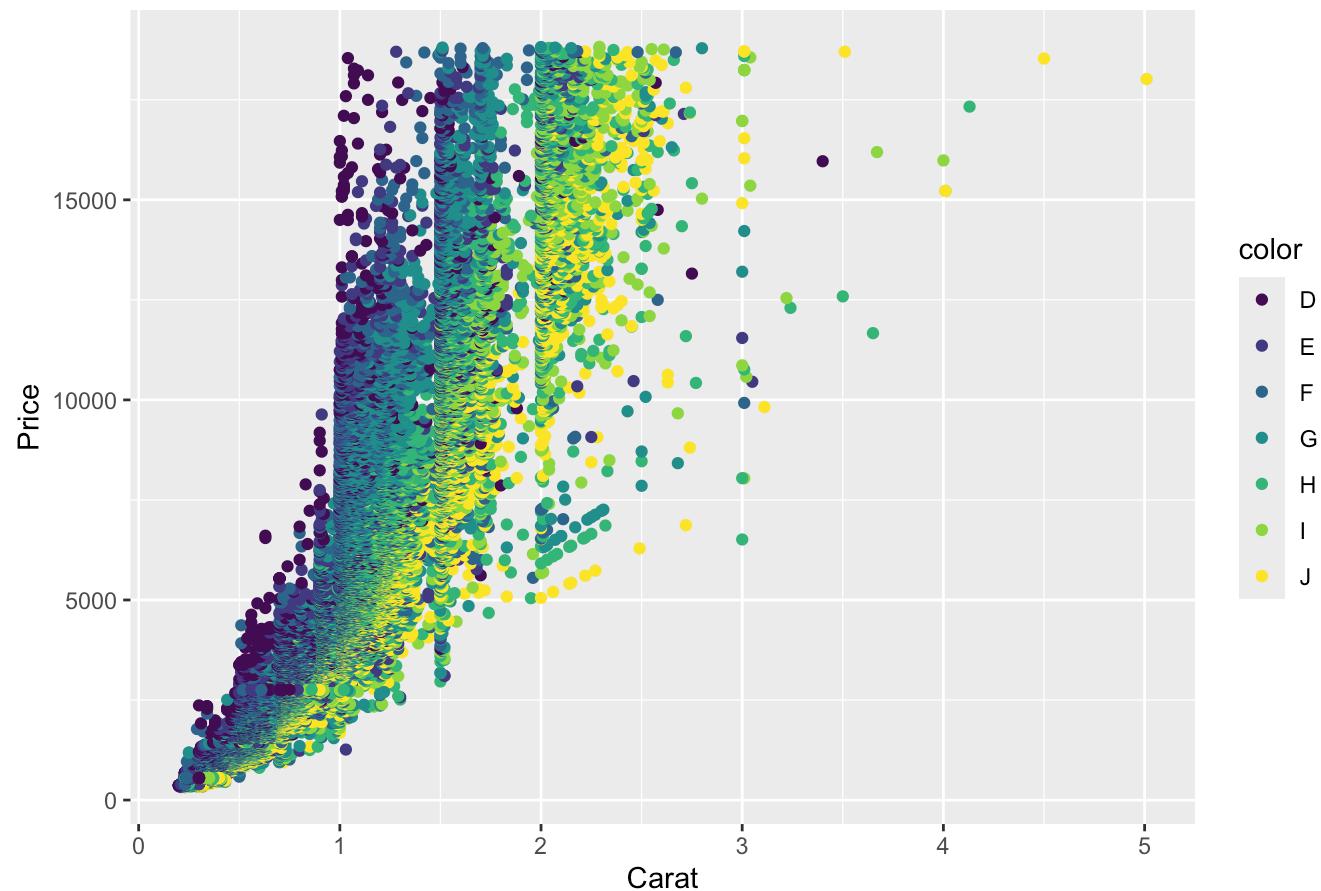
## 2.9 Deliverable 11: Re-Build Your Scatterplot with Color

► [Code](#)



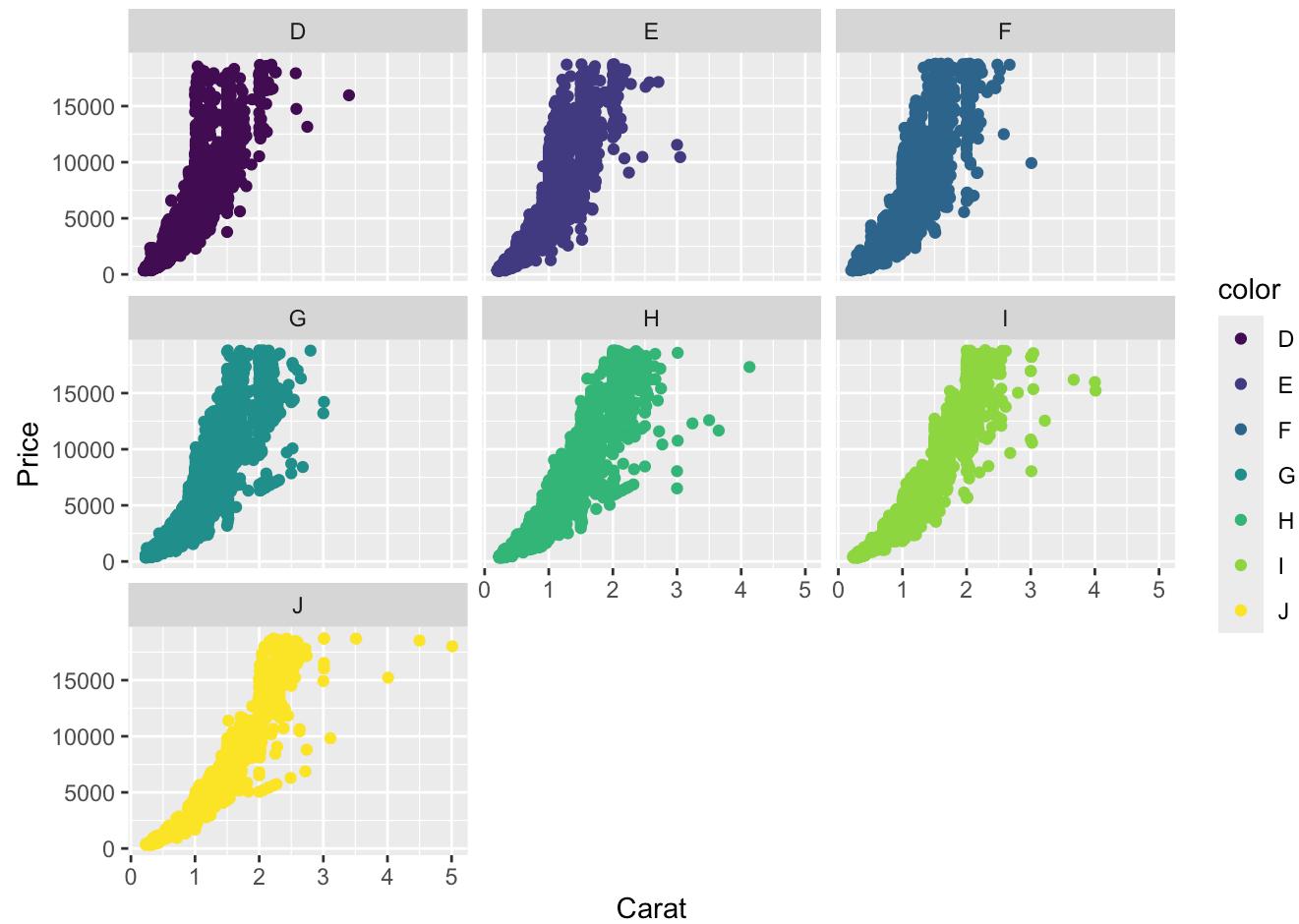
► Code

### Color Scatterplot of Carat vs Price



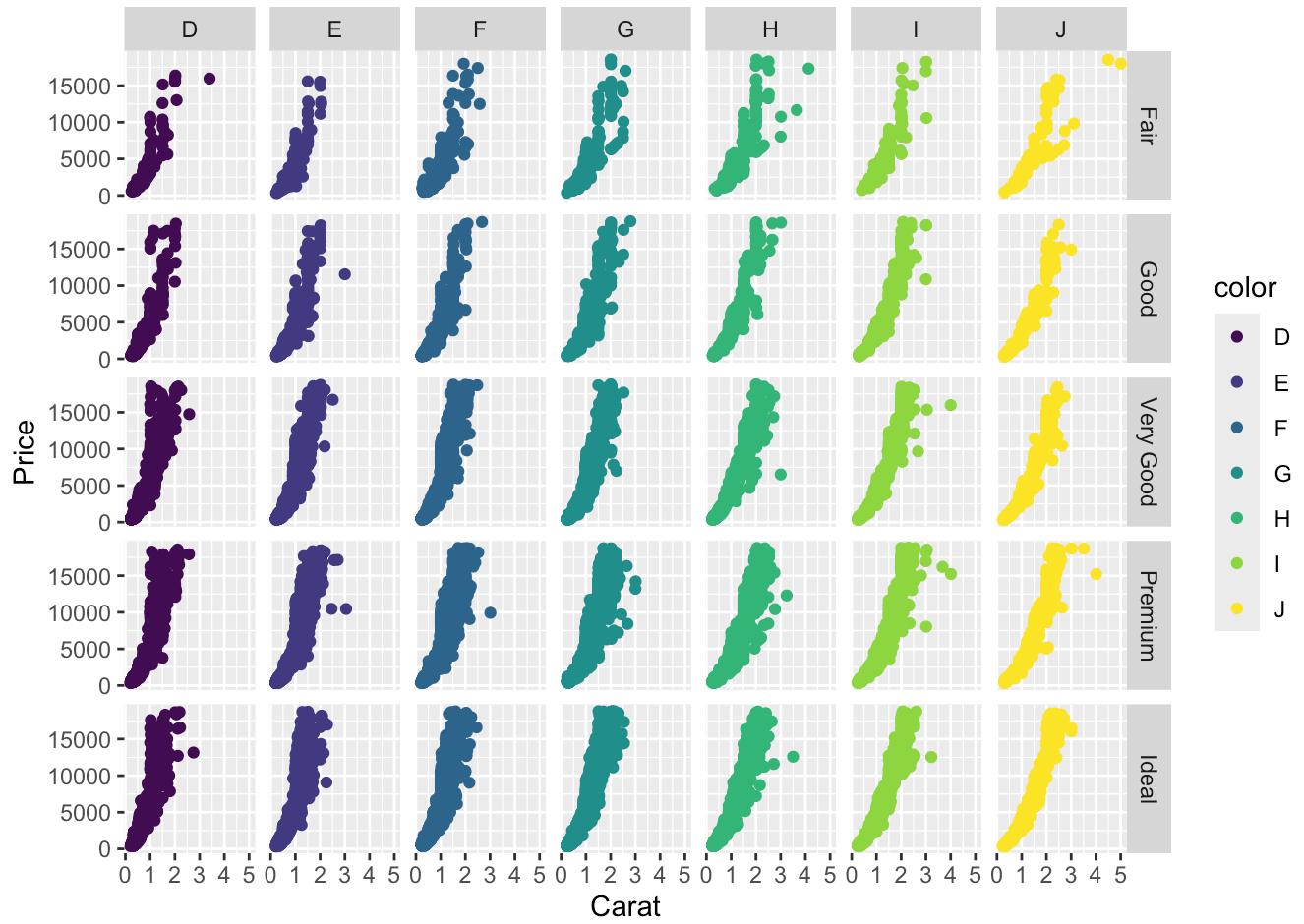
## 2.10 Deliverable 12: Demonstrate a Facet Wrap Plot

► [Code](#)



## 2.11 Deliverable 13: Demonstrate a Facet Wrap Grid Plot

► Code

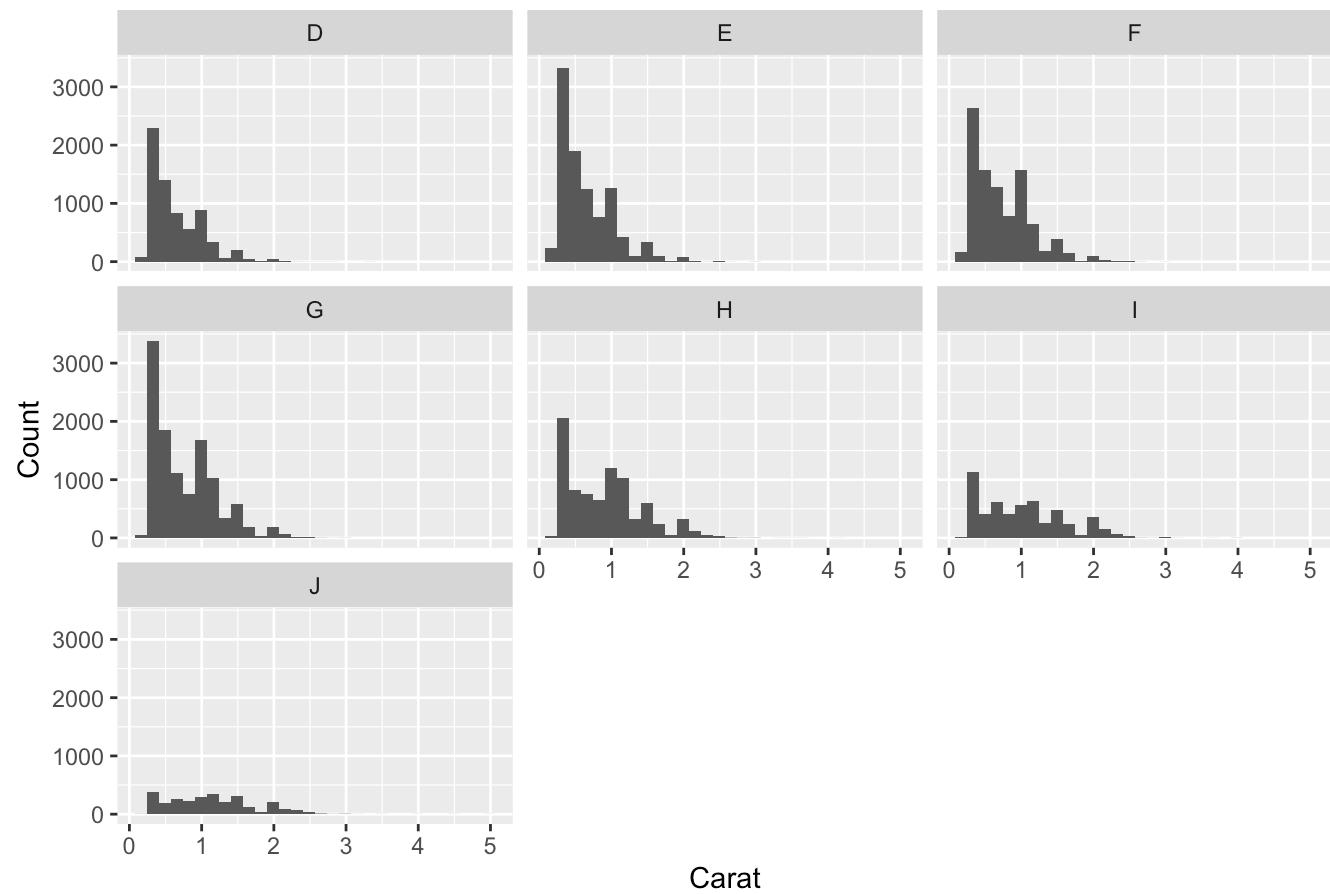


## 2.12 Deliverable 14: Demonste a Facet Wrap Plot with a Histogram

► Code

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

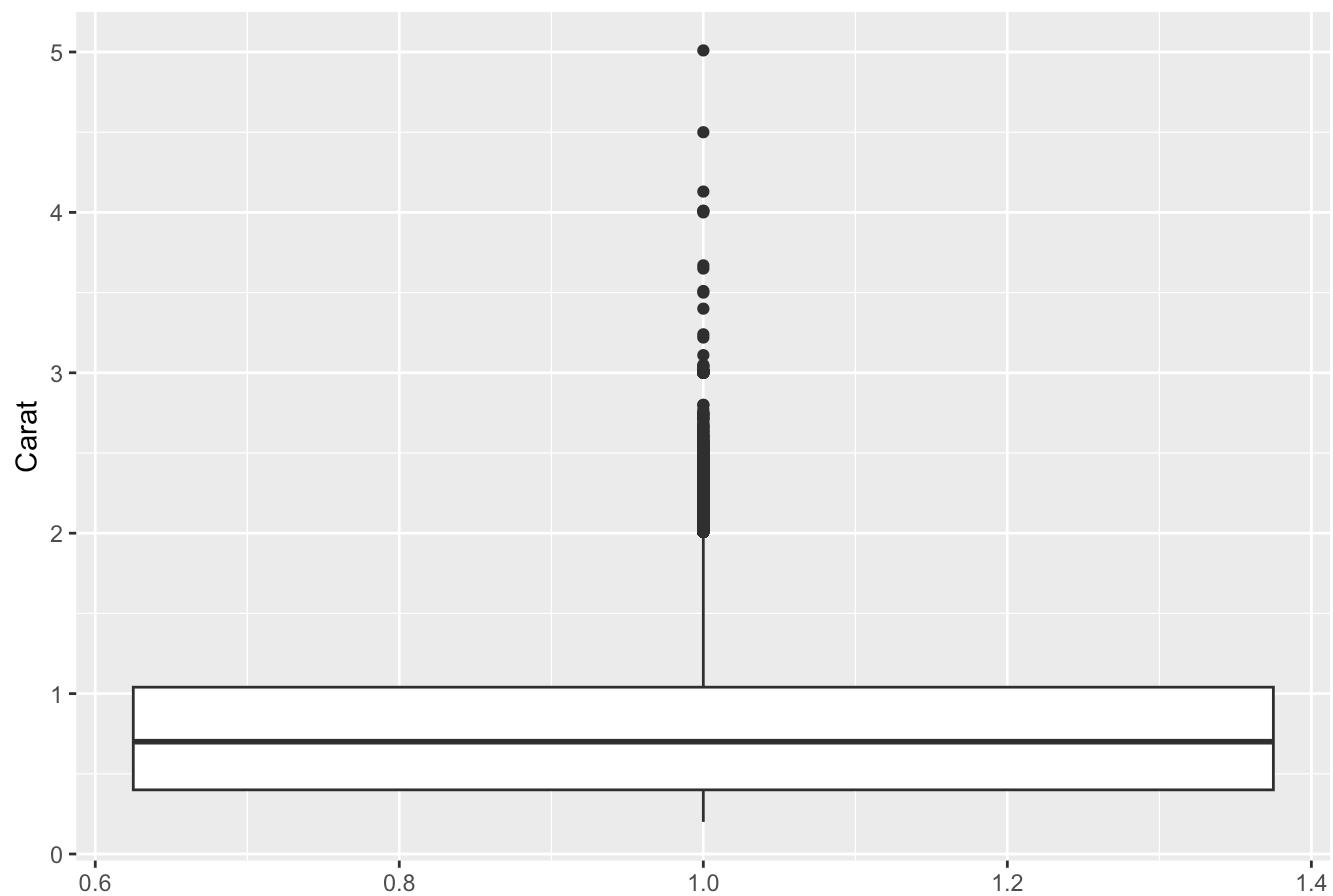
### Histogram of Carat by Color



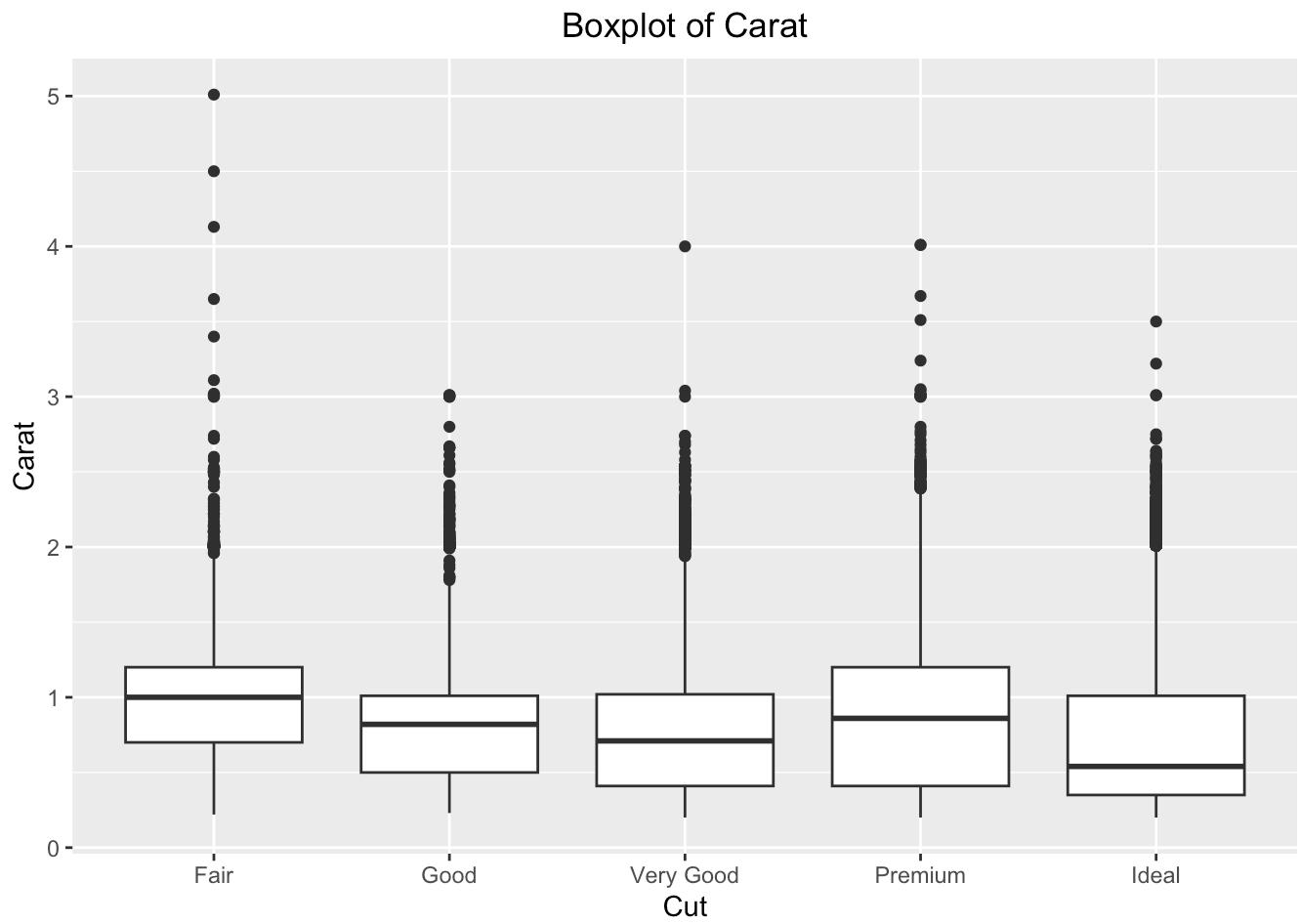
## 2.13 Deliverable 15: Demonstrate Boxplots in ggplot

► Code

Boxplot of Carat

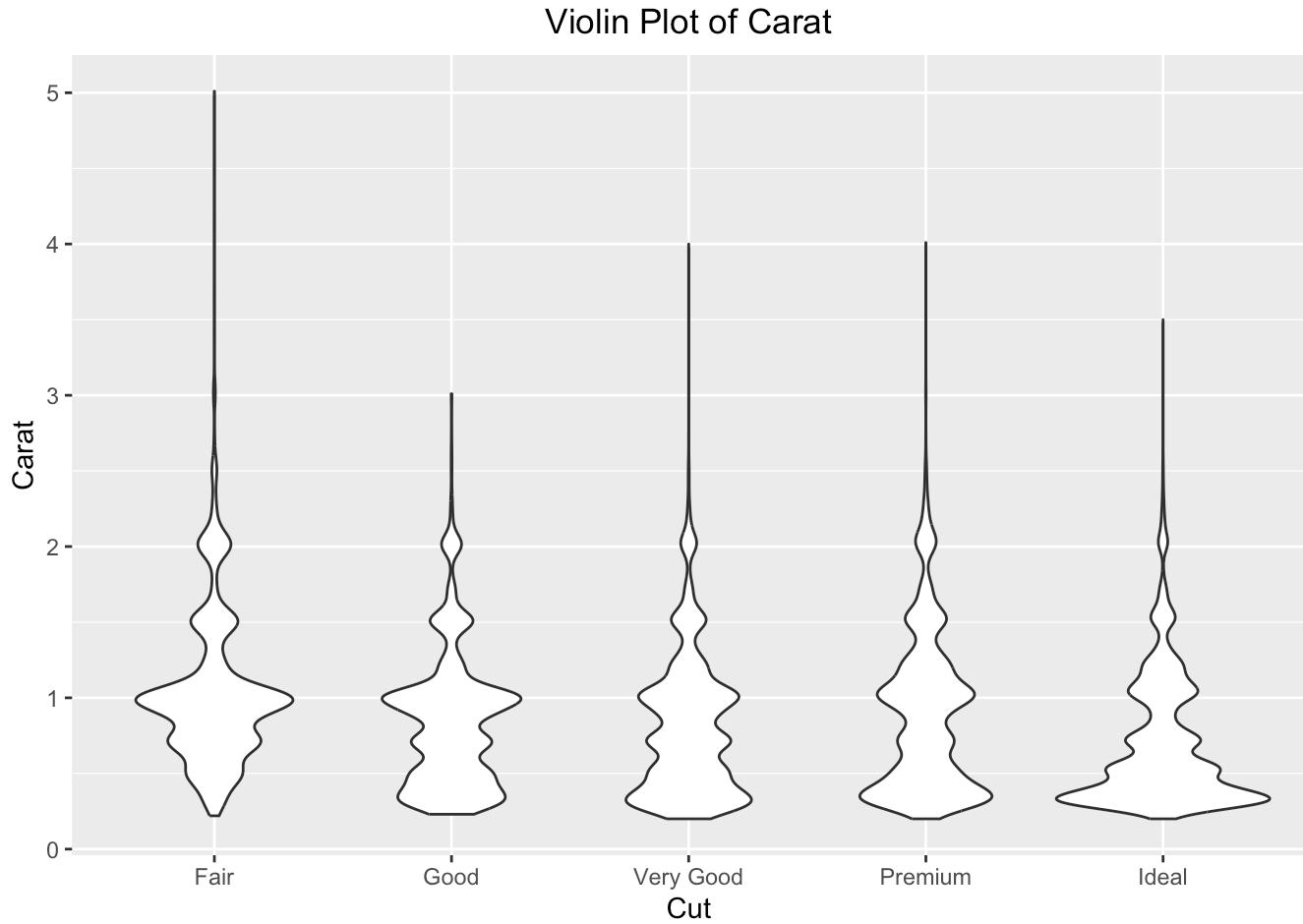


► Code



## 2.14 Deliverable 16: Demonstrate a Violinplot in ggplot

► Code



## 3 Part 2: Reading in External Data and Visualizing with ggplot2

### 3.1 Deliverable 17: Read in a CSV file from the Web, Explore, and Summarize

► Code

```
Rows: 1269 Columns: 17
— Column specification —
Delimiter: ","
chr (8): name, city, state, region, highest_degree, control, gender, loan_de...
dbl (9): id, admission_rate, sat_avg, undergrads, tuition, faculty_salary_av...
i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

► Code

```
# A tibble: 6 × 17
  id name    city   state region highest_degree control gender admission_rate
  <dbl> <chr>   <chr> <chr> <chr>       <chr>   <chr>      <dbl>
1 102669 Alaska... Anch... AK     West   Graduate   Private CoEd    0.421
2 101648 Marion... Mari... AL     South  Associate  Public  CoEd    0.614
3 100830 Auburn... Mont... AL     South  Graduate   Public  CoEd    0.802
4 101879 Univer... Flor... AL     South  Graduate   Public  CoEd    0.679
5 100858 Auburn... Aubu... AL     South  Graduate   Public  CoEd    0.835
6 100663 Univer... Birm... AL     South  Graduate   Public  CoEd    0.857
# i 8 more variables: sat_avg <dbl>, undergrads <dbl>, tuition <dbl>,
# faculty_salary_avg <dbl>, loan_default_rate <chr>, median_debt <dbl>,
# lon <dbl>, lat <dbl>
```

► Code

```
# A tibble: 6 × 17
  id name    city   state region highest_degree control gender admission_rate
  <dbl> <chr>   <chr> <chr> <chr>       <chr>   <chr>      <dbl>
1 237969 West V... Buck... WV     South  Graduate   Private CoEd    0.776
2 237385 Glenvi... Glen... WV     South  Bachelor  Public  CoEd    0.466
3 237118 Alders... Phil... WV     South  Graduate   Private CoEd    0.426
4 238078 Wheeli... Whee... WV     South  Graduate   Private CoEd    0.626
5 237181 Bethan... Beth... WV     South  Graduate   Private CoEd    0.620
6 240727 Univer... Lara... WY     West   Graduate   Public  CoEd    0.977
# i 8 more variables: sat_avg <dbl>, undergrads <dbl>, tuition <dbl>,
# faculty_salary_avg <dbl>, loan_default_rate <chr>, median_debt <dbl>,
# lon <dbl>, lat <dbl>
```

► Code

```
[1] "spec_tbl_df" "tbl_df"        "tbl"           "data.frame"
```

► Code

	id	name	city	state
Min.	:100654	Length:1269	Length:1269	Length:1269
1st Qu.	:153250	Class :character	Class :character	Class :character
Median	:186283	Mode :character	Mode :character	Mode :character
Mean	:186988			
3rd Qu.	:215284			
Max.	:484905			
	region	highest_degree	control	gender
Length:	1269	Length:1269	Length:1269	Length:1269
Class :	character	Class :character	Class :character	Class :character
Mode :	character	Mode :character	Mode :character	Mode :character

	admission_rate	sat_avg	undergrads	tuition
Min.	:0.0509	Min. : 720	Min. : 47	Min. : 2732
1st Qu.	:0.5339	1st Qu.: 973	1st Qu.: 1296	1st Qu.: 8970

```

Median :0.6687  Median :1040   Median : 2556  Median :20000
Mean    :0.6501  Mean   :1060   Mean    : 5629  Mean    :21025
3rd Qu.:0.7859  3rd Qu.:1120   3rd Qu.: 6715  3rd Qu.:30364
Max.   :1.0000   Max.   :1545   Max.   :52280   Max.   :51008
faculty_salary_avg loan_default_rate median_debt          lon
Min.   : 1451    Length:1269      Min.   : 6056  Min.   :-157.92
1st Qu.: 6191    Class :character  1st Qu.:21250 1st Qu.: -94.17
Median : 7272    Mode   :character  Median :24588  Median : -84.89
Mean   : 7656    Mode   :character  Mean   :23483  Mean   : -88.29
3rd Qu.: 8671    Mode   :character  3rd Qu.:27000 3rd Qu.: -78.63
Max.   :20650    Mode   :character  Max.   :41000  Max.   : -68.59
lat
Min.   :19.71
1st Qu.:35.22
Median :39.74
Mean   :38.61
3rd Qu.:41.81
Max.   :61.22

```

## 3.2 Deliverable 18: Wrangle the Data and Re-Summarize

► Code

```

id           name          city          state
Min.   :100654  Length:1269      Length:1269      PA    :101
1st Qu.:153250  Class :character  Class :character  NY    : 84
Median :186283  Mode   :character  Mode   :character  CA    : 71
Mean   :186988
3rd Qu.:215284
Max.   :484905
region      highest_degree control          gender
Midwest   :353   Associate: 20  Length:1269      Length:1269
Northeast :299   Bachelor : 200  Class :character  Class :character
South     :459   Graduate :1049  Mode   :character  Mode   :character
West      :158

```

```

admission_rate      sat_avg      undergrads      tuition
Min.   :0.0509  Min.   : 720  Min.   : 47  Min.   : 2732
1st Qu.:0.5339  1st Qu.: 973  1st Qu.:1296  1st Qu.: 8970
Median :0.6687  Median :1040  Median : 2556  Median :20000
Mean   :0.6501  Mean   :1060  Mean   : 5629  Mean   :21025
3rd Qu.:0.7859  3rd Qu.:1120  3rd Qu.: 6715  3rd Qu.:30364
Max.   :1.0000   Max.   :1545  Max.   :52280   Max.   :51008

faculty_salary_avg loan_default_rate median_debt          lon
Min.   : 1451    Length:1269      Min.   : 6056  Min.   :-157.92

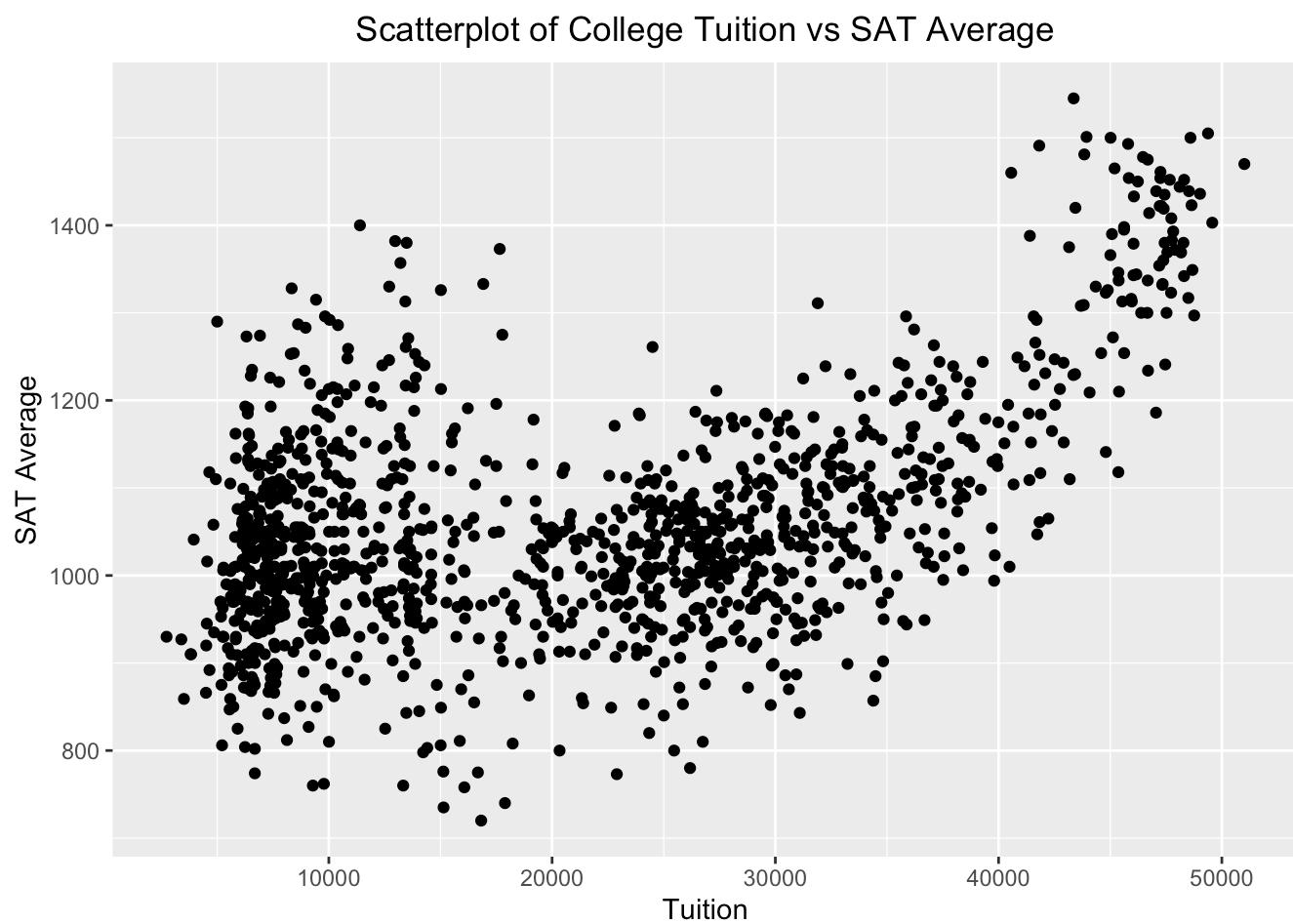
```

```
1st Qu.: 6191      Class :character  1st Qu.:21250   1st Qu.: -94.17  
Median : 7272      Mode  :character  Median :24588    Median : -84.89  
Mean   : 7656          Mean   :23483    Mean   : -88.29  
3rd Qu.: 8671      3rd Qu.:27000   3rd Qu.: -78.63  
Max.   :20650      Max.   :41000    Max.   : -68.59
```

```
lat  
Min.  :19.71  
1st Qu.:35.22  
Median :39.74  
Mean   :38.61  
3rd Qu.:41.81  
Max.   :61.22
```

### 3.3 Deliverable 19: Create a Basic Scatterplot of the college Data

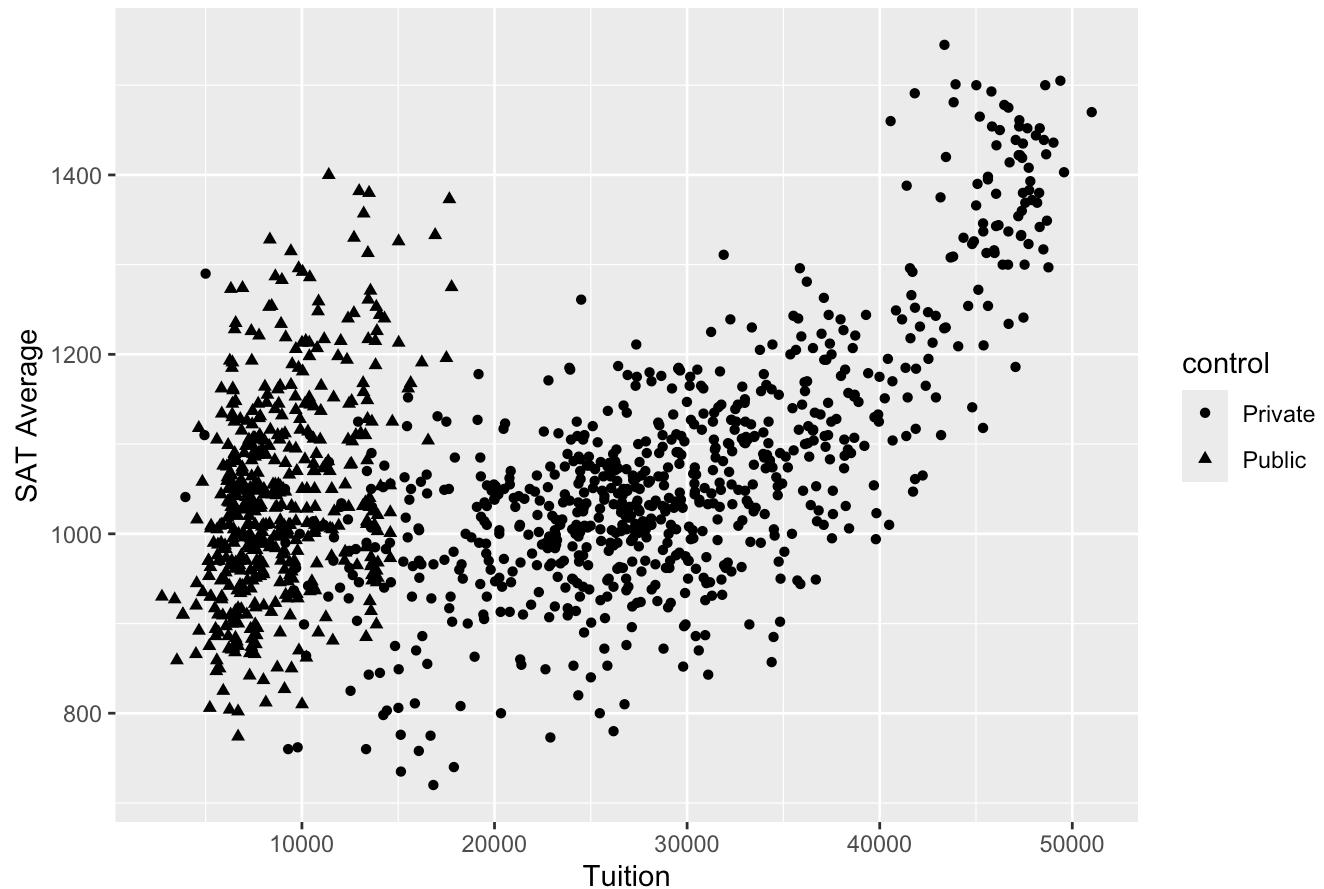
► Code



### 3.4 Deliverable 20: Add Shape Control to Your Scatterplot

► Code

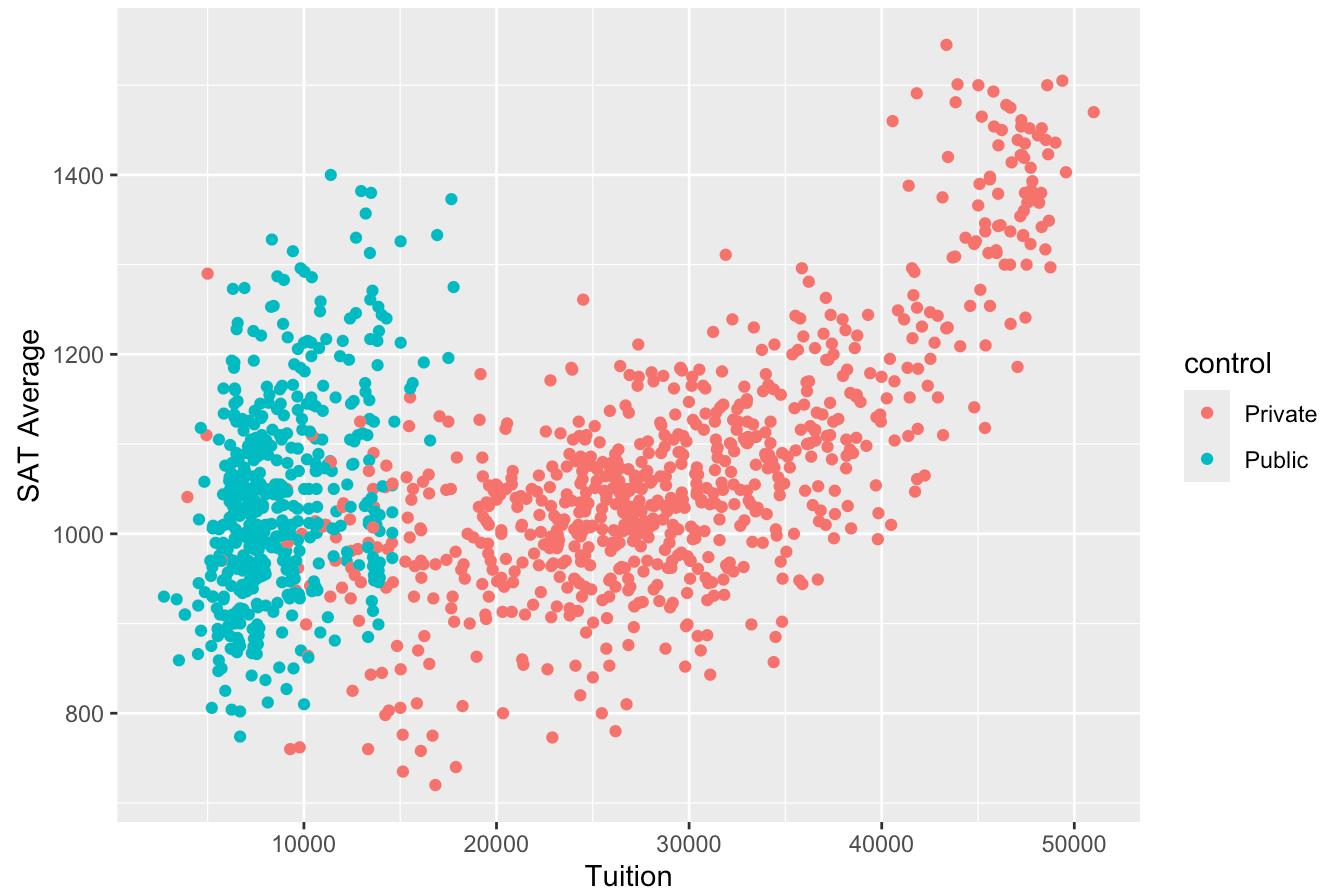
Scatterplot of College Tuition vs SAT Average by Control



### 3.5 Deliverable 21: Add Color to Your Shape Control

► Code

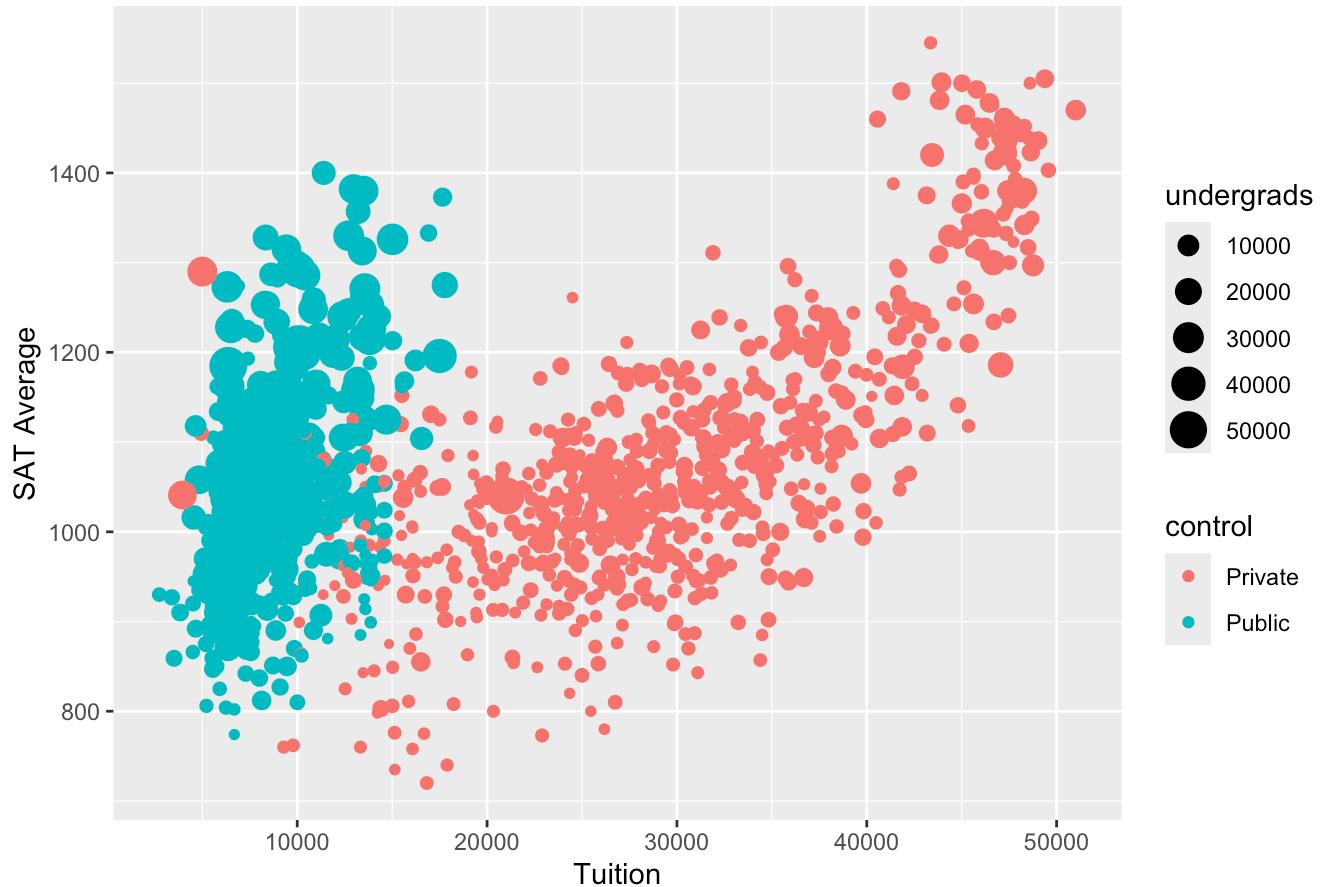
Scatterplot of College Tuition vs SAT Average by Control



### 3.6 Deliverable 22: Add Size to Your AES and Interpret

► Code

Scatterplot of College Tuition vs SAT Average by Control



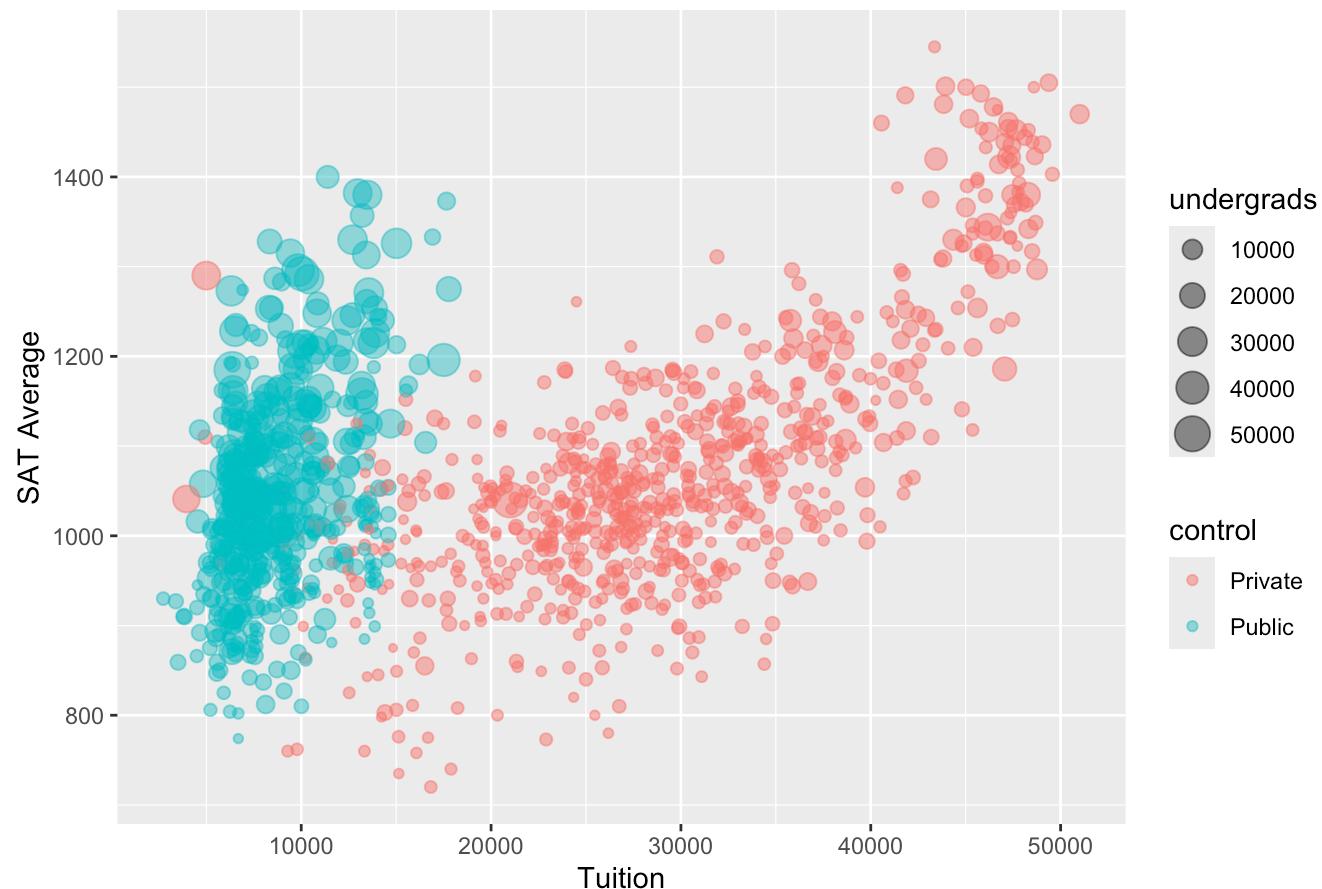
### 3.6.0.1 Analysis:

There appears to be an inverse relationship between SAT Score and Tuition. The concentration of larger bubbles in the top left corner (intersection of increased SAT Scores and decreased tuition) and the smaller bubbles concentrated in the lower right region (intersection of decreased SAT Scores and increased tuition) support this from visual analysis.

## 3.7 Deliverable 23: Add Alpha (Opacity) to Your AES

► Code

Scatterplot of College Tuition vs SAT Average by Control

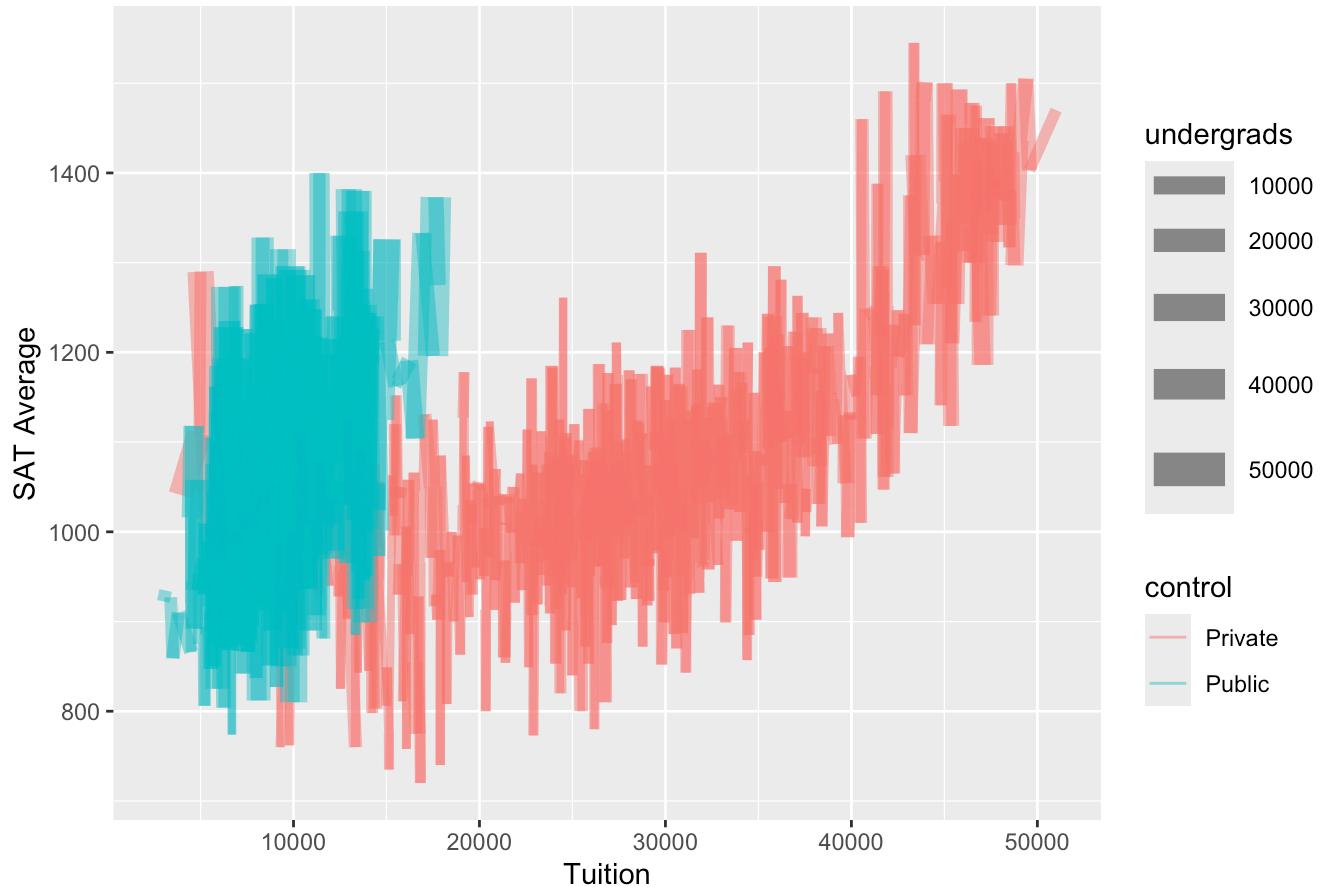


### 3.8 Deliverable 24: Add a Line Plot and Smooth

► Code

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
i Please use `linewidth` instead.
```

## Scatterplot of College Tuition vs SAT Average by Control

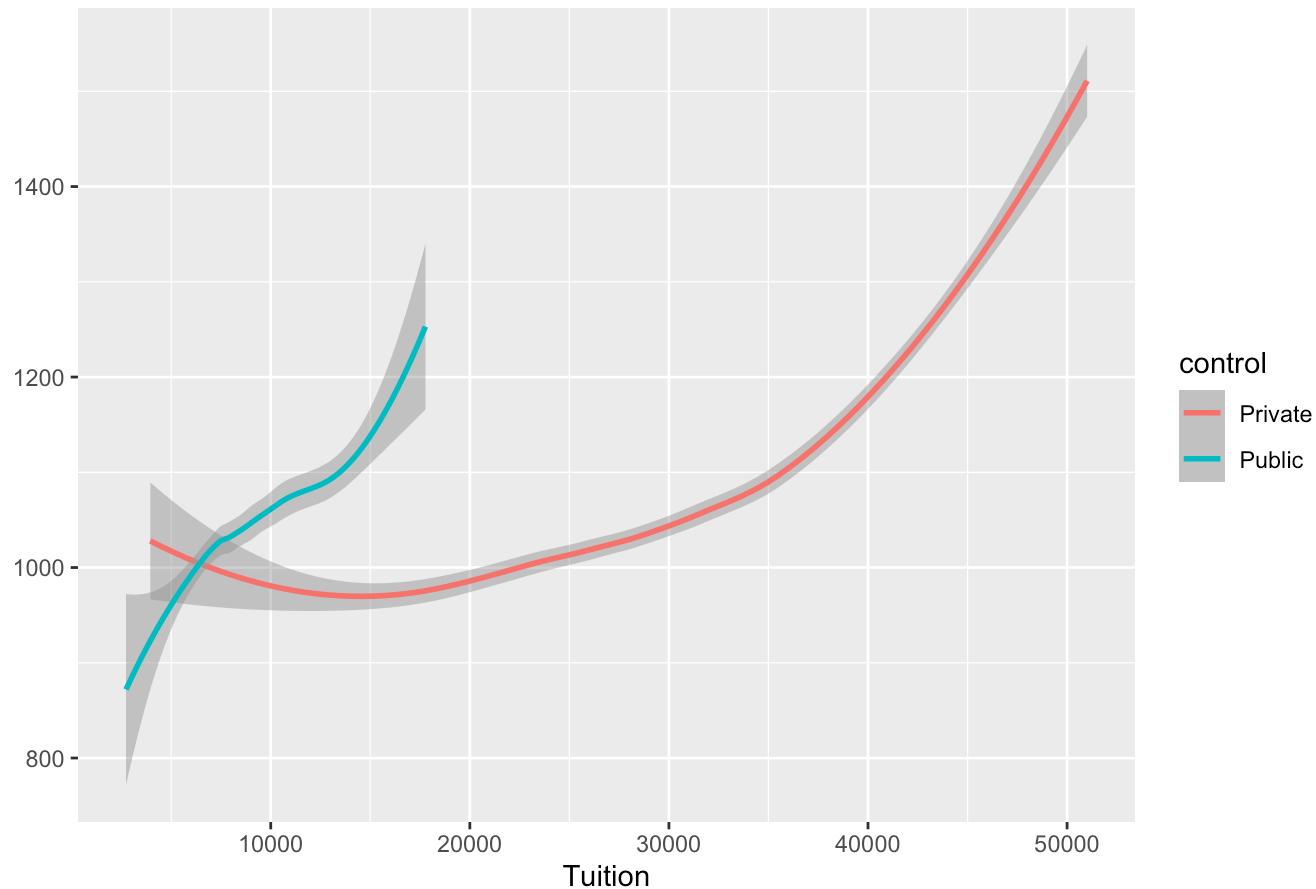


### ► Code

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

Warning: The following aesthetics were dropped during statistical transformation: size.  
i This can happen when ggplot fails to infer the correct grouping structure in  
the data.  
i Did you forget to specify a `group` aesthetic or to convert a numerical  
variable into a factor?

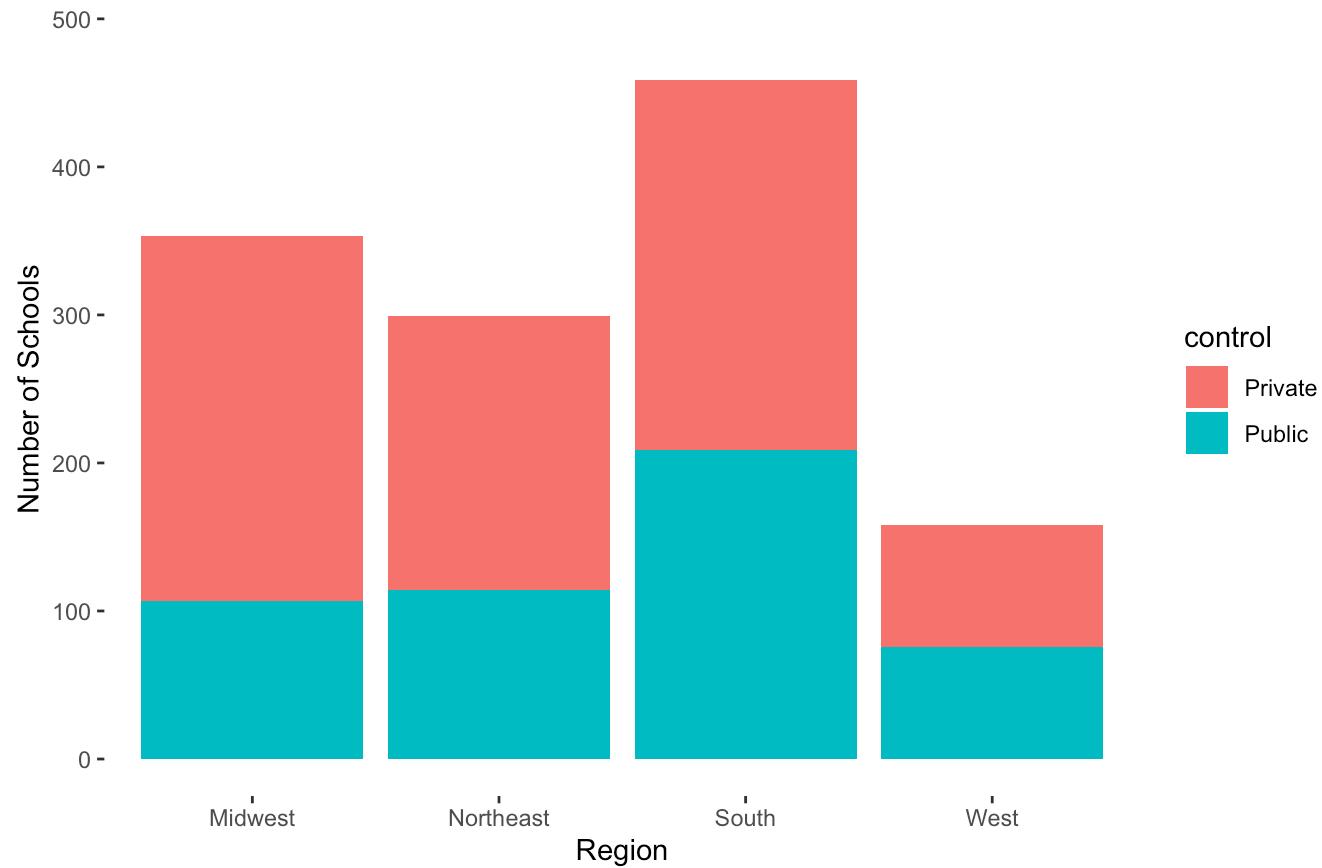
Line Chart of College Tuition vs SAT Average by Control



### 3.9 Deliverable 25: Create a Bar Graph

► Code

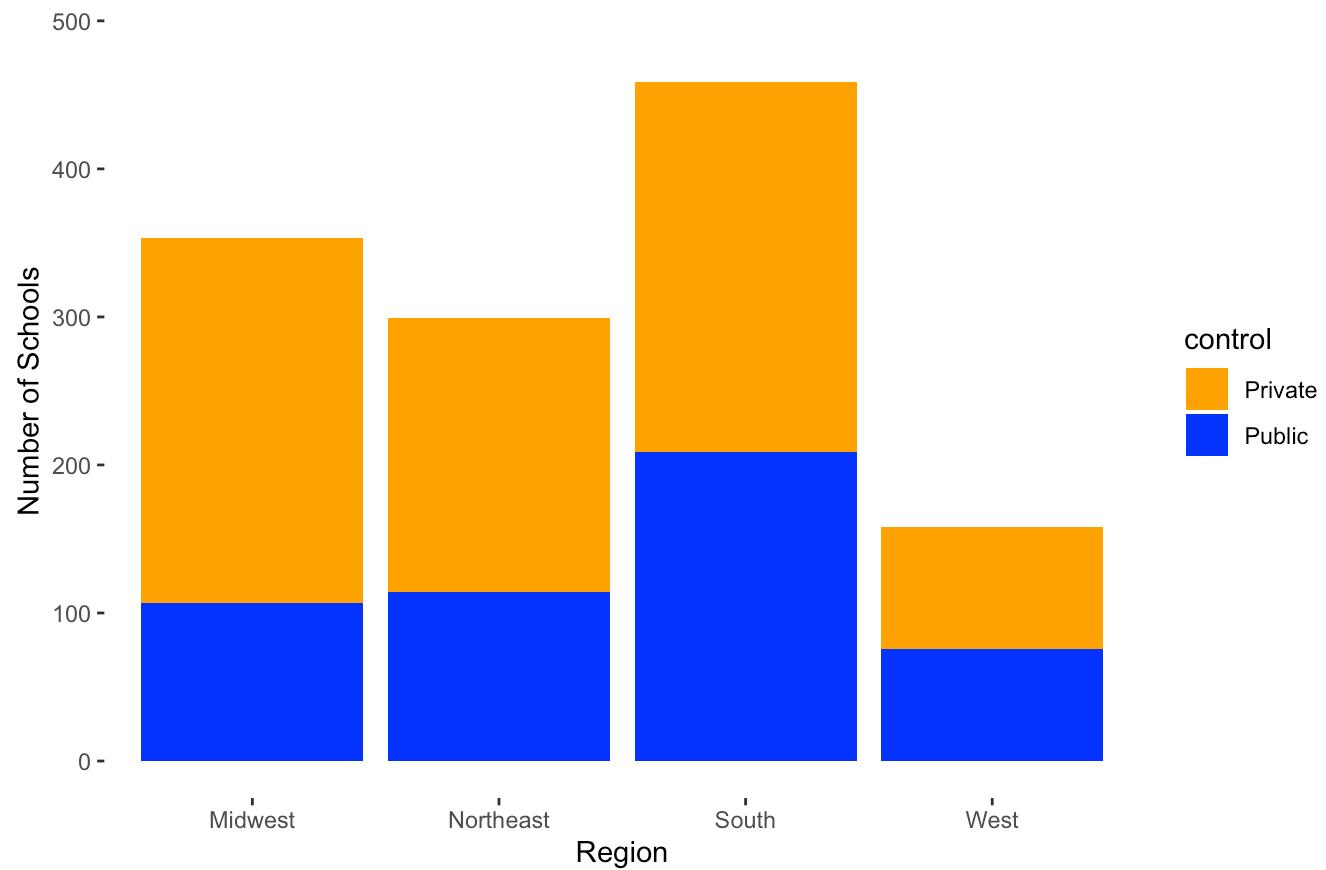
Bar Graph of Number of Schools by Region and Control Type



### 3.10 Deliverable 26: Change Colors on your Bar Graph

► Code

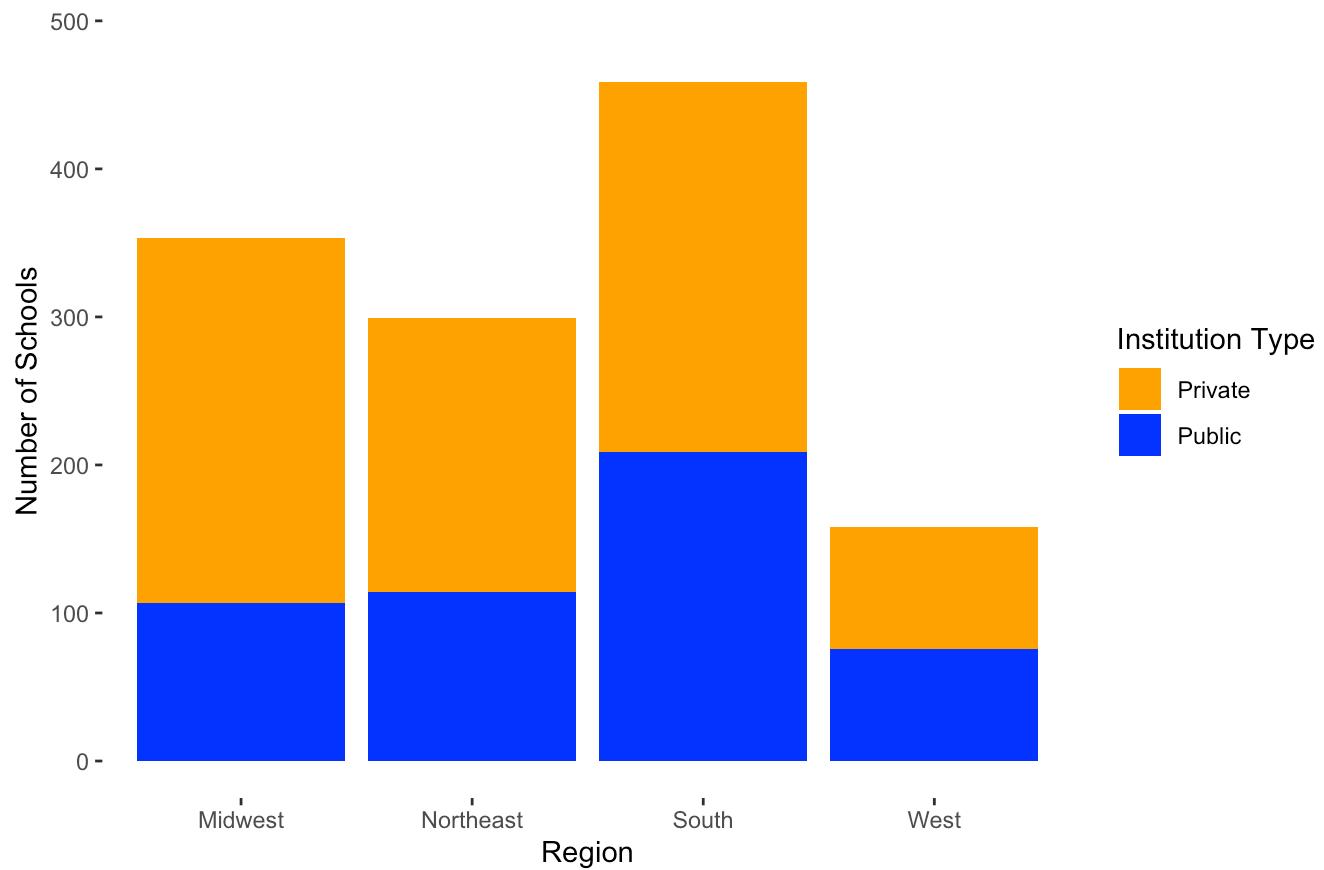
## Bar Graph of Number of Schools by Region and Control Type



### 3.11 Deliverable 27: Add Legends and Titles to Your Bar Graphs

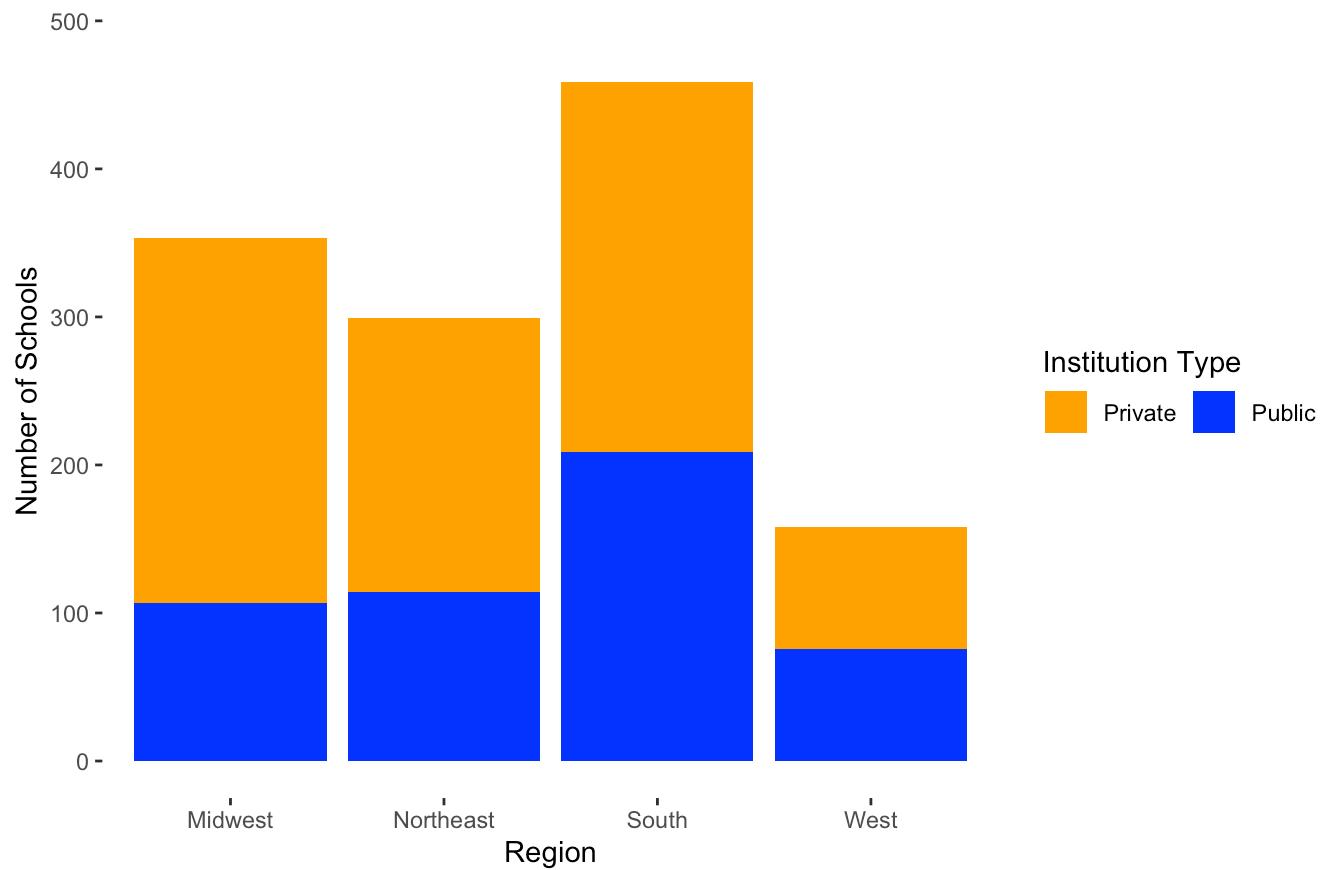
► Code

## Bar Graph of Number of Schools by Region and Control Type



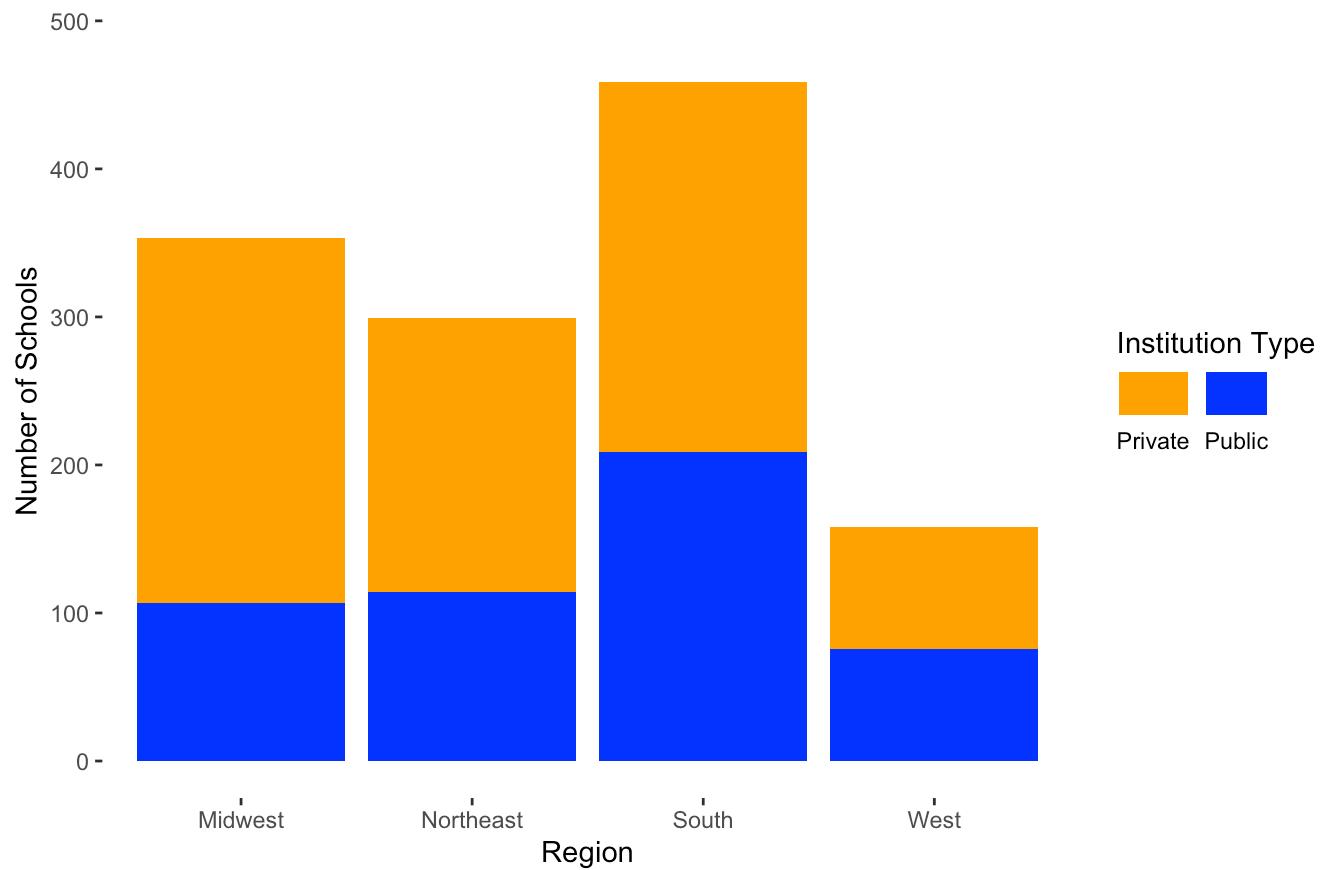
► Code

## Bar Graph of Number of Schools by Region and Control Type



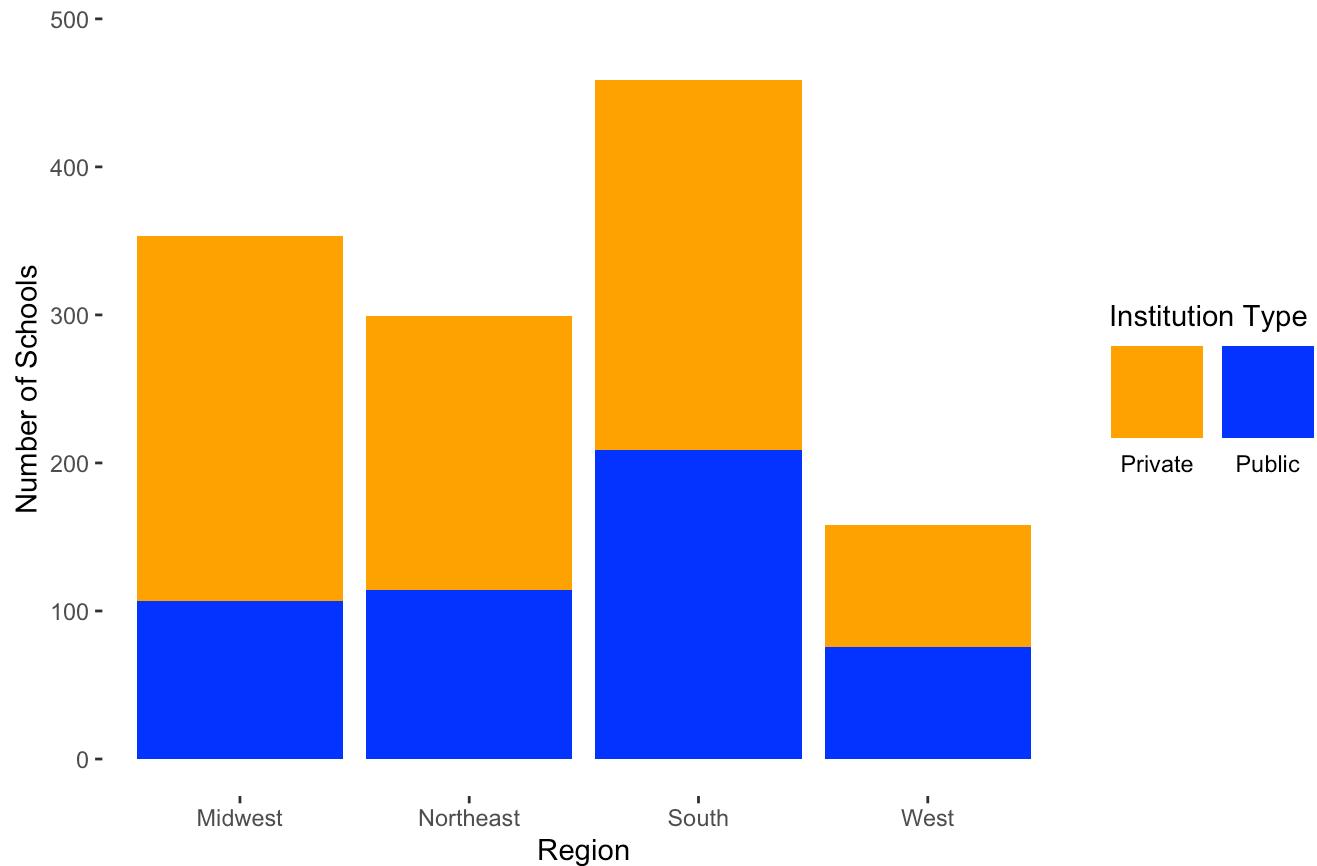
► Code

## Bar Graph of Number of Schools by Region and Control Type



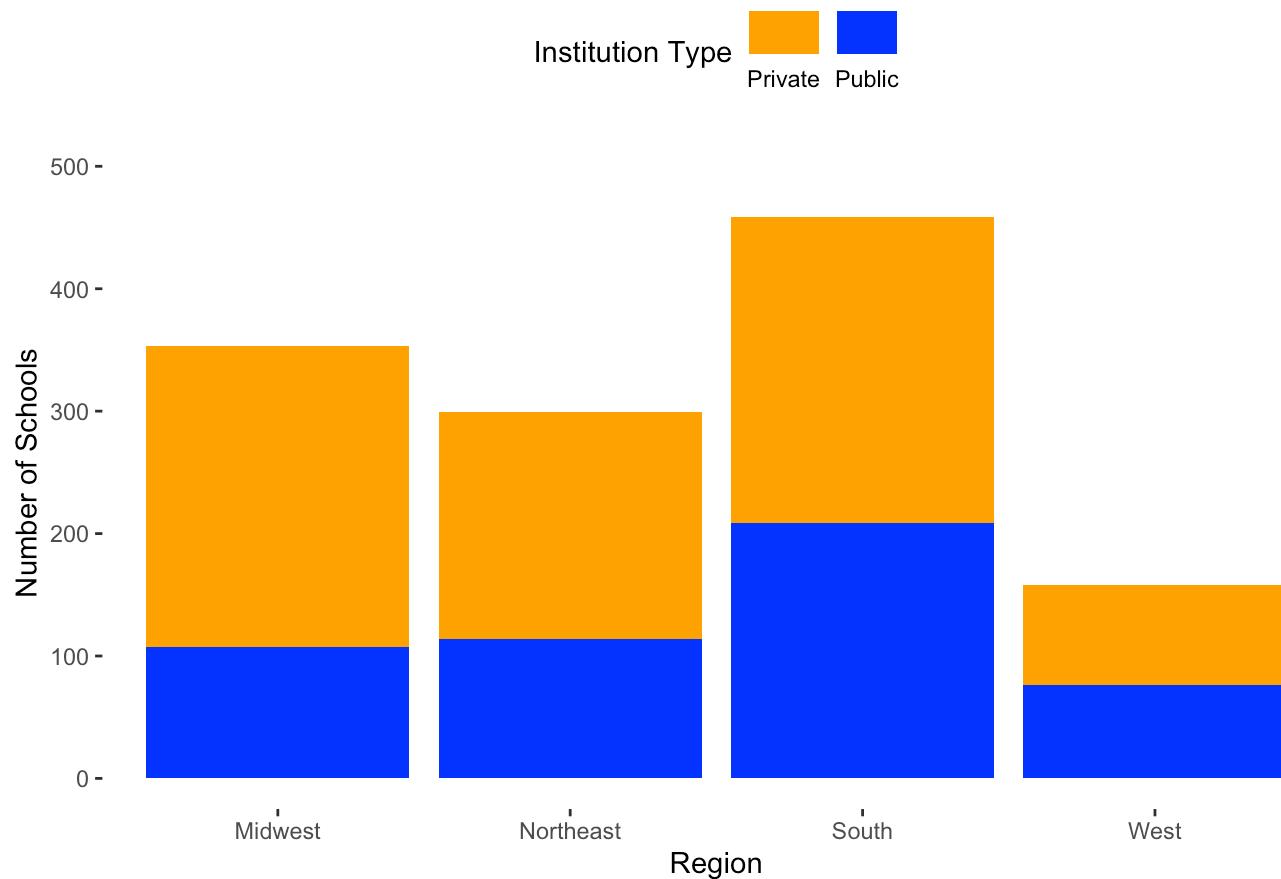
► Code

## Bar Graph of Number of Schools by Region and Control Type



► Code

## Bar Graph of Number of Schools by Region and Control Type

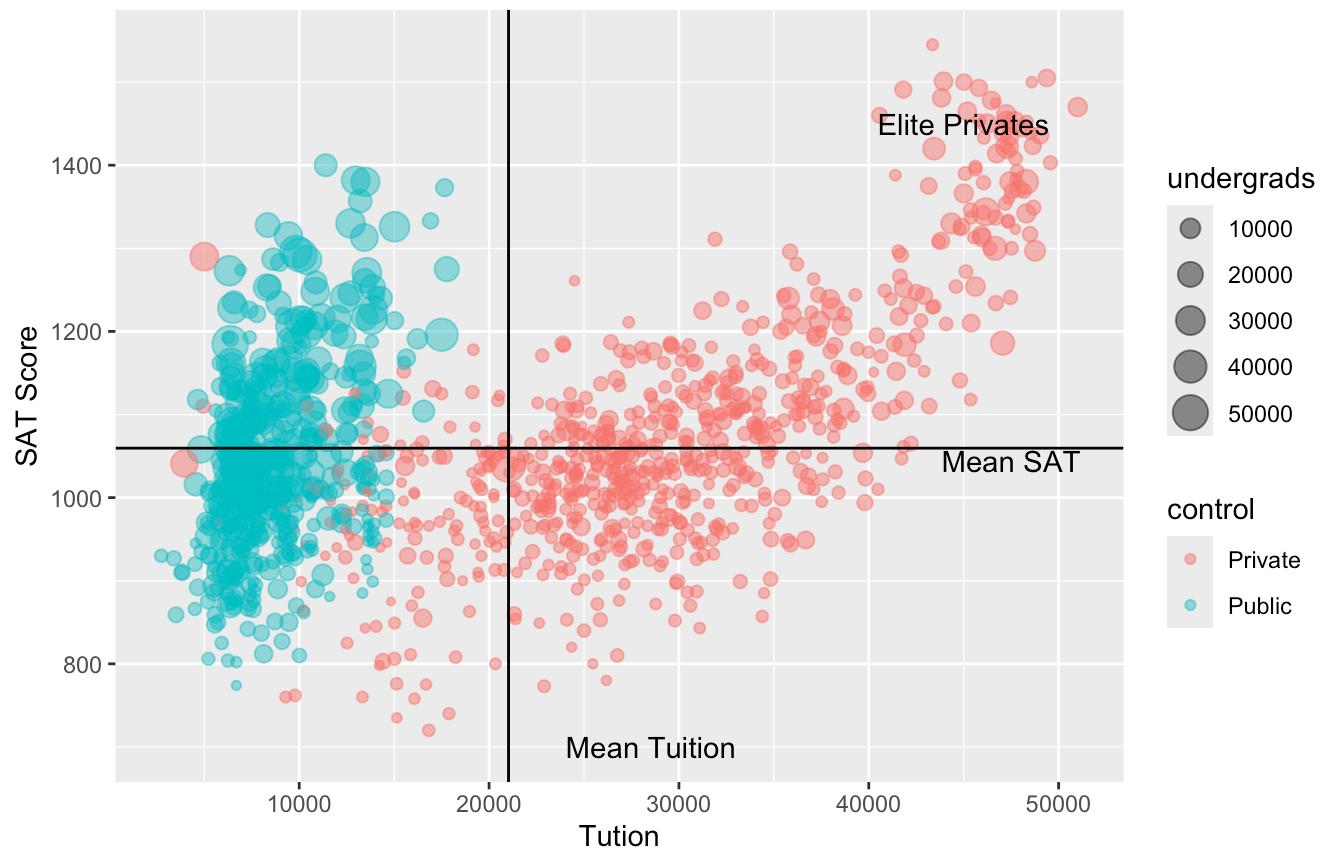


## 3.12 Deliverable 28: Add Annotations to Your Bar Graphs

► Code

## College Data: SAT Scores and Tuition by Control Type

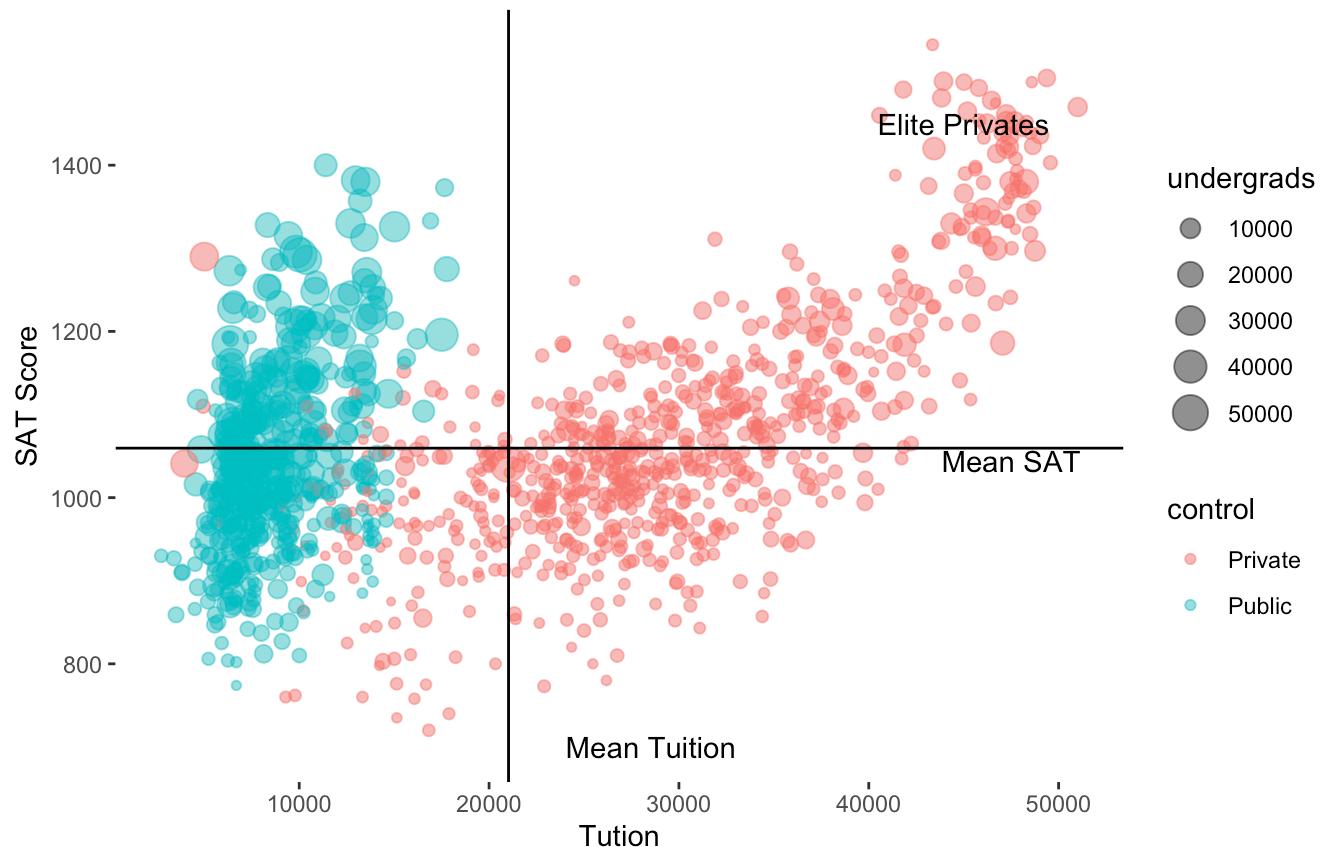
Size of circle represents number of undergraduates



► Code

## College Data: SAT Scores and Tuition by Control Type

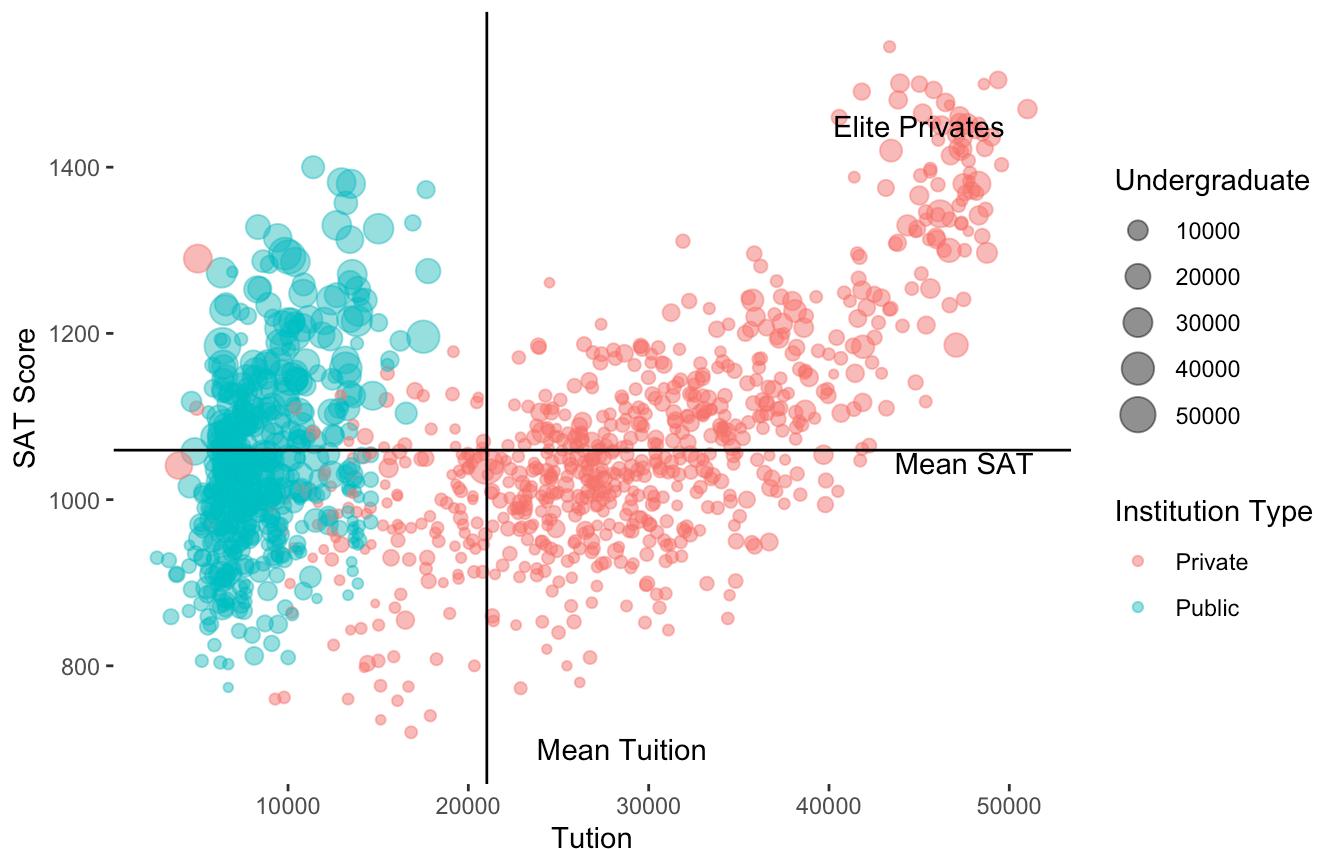
Size of circle represents number of undergraduates



► Code

## College Data: SAT Scores and Tuition by Control Type

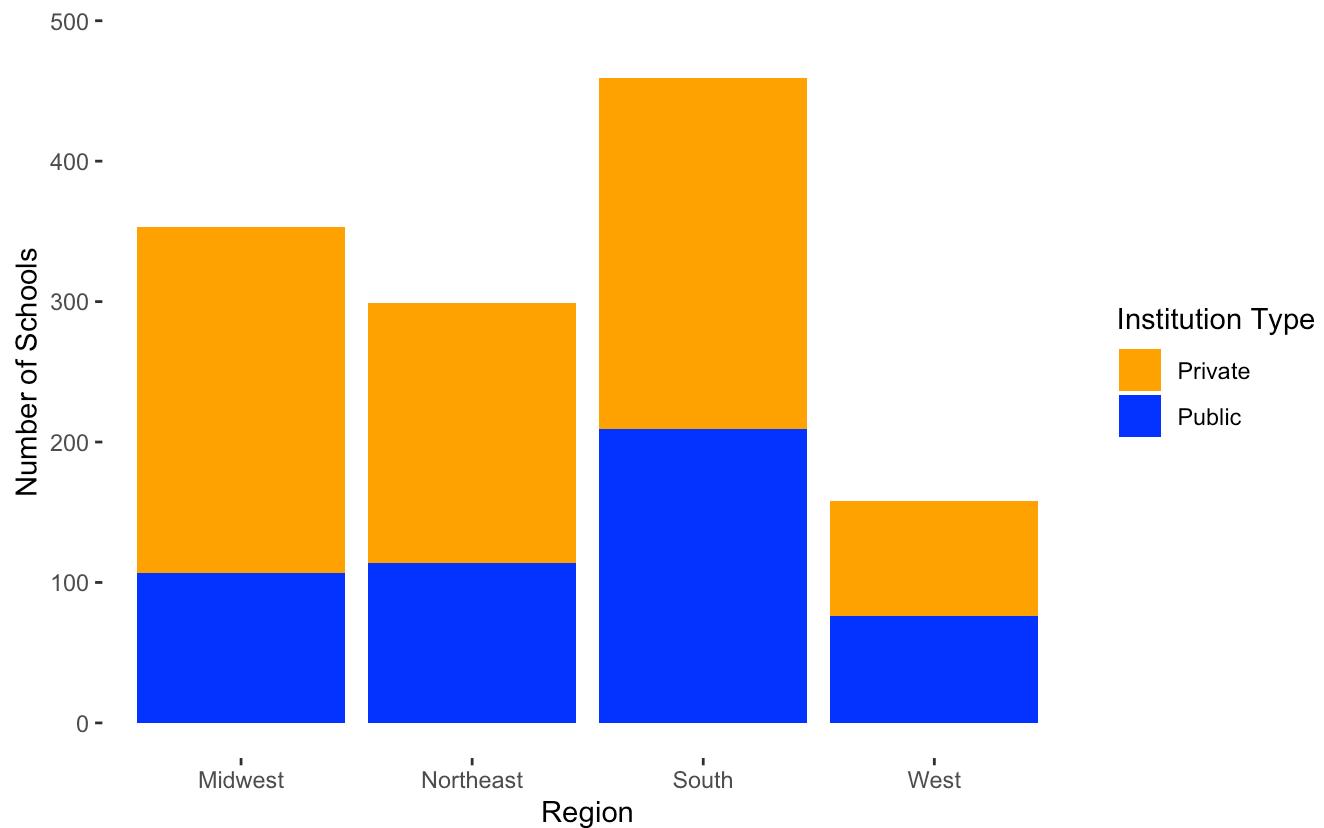
Size of circle represents number of undergraduates



► Code

More college are in the Southern USA than any other region.

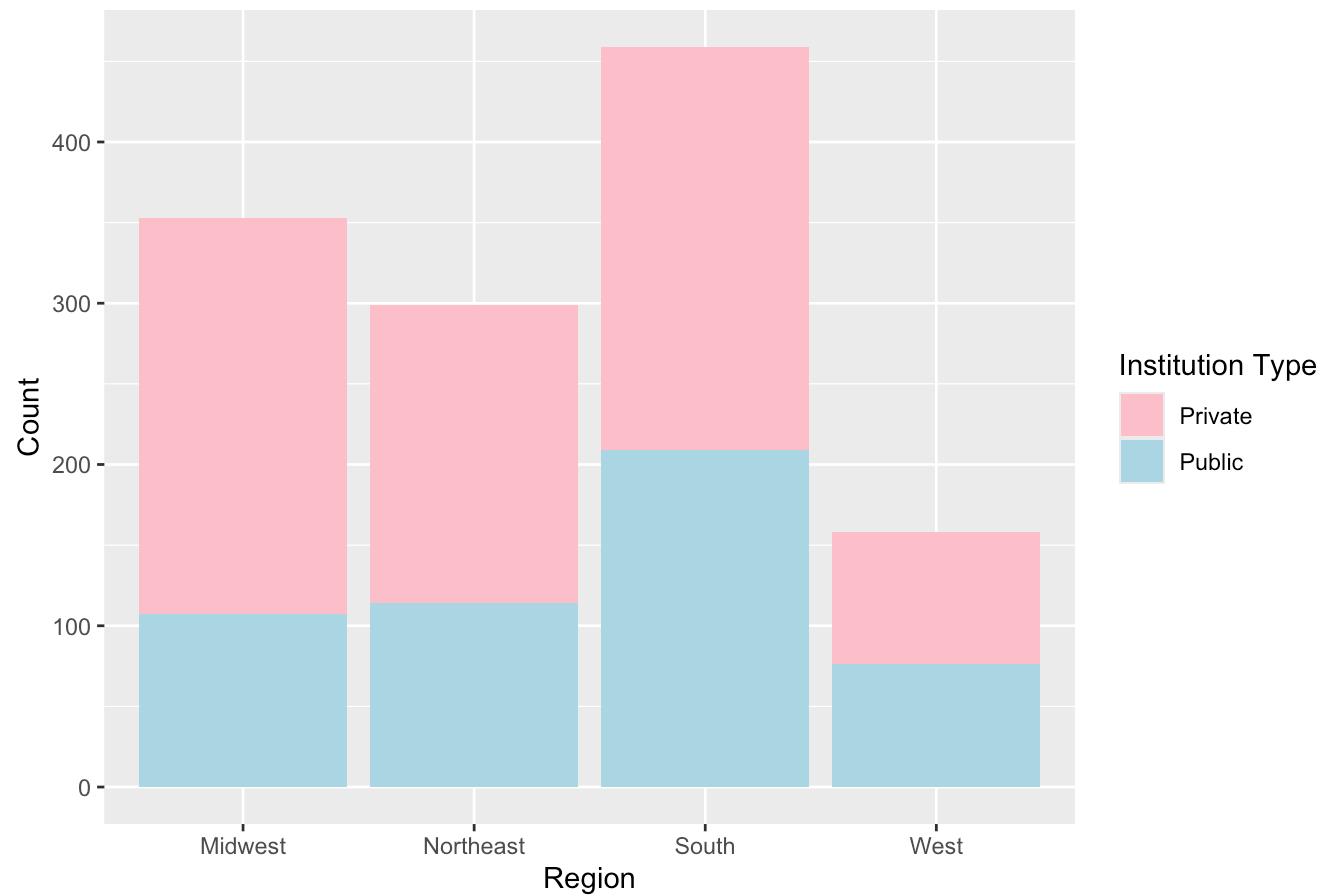
Source: Collegeboard



### 3.13 Deliverable 29: Explore Pre-Defined Themes for Your Bar Graphs

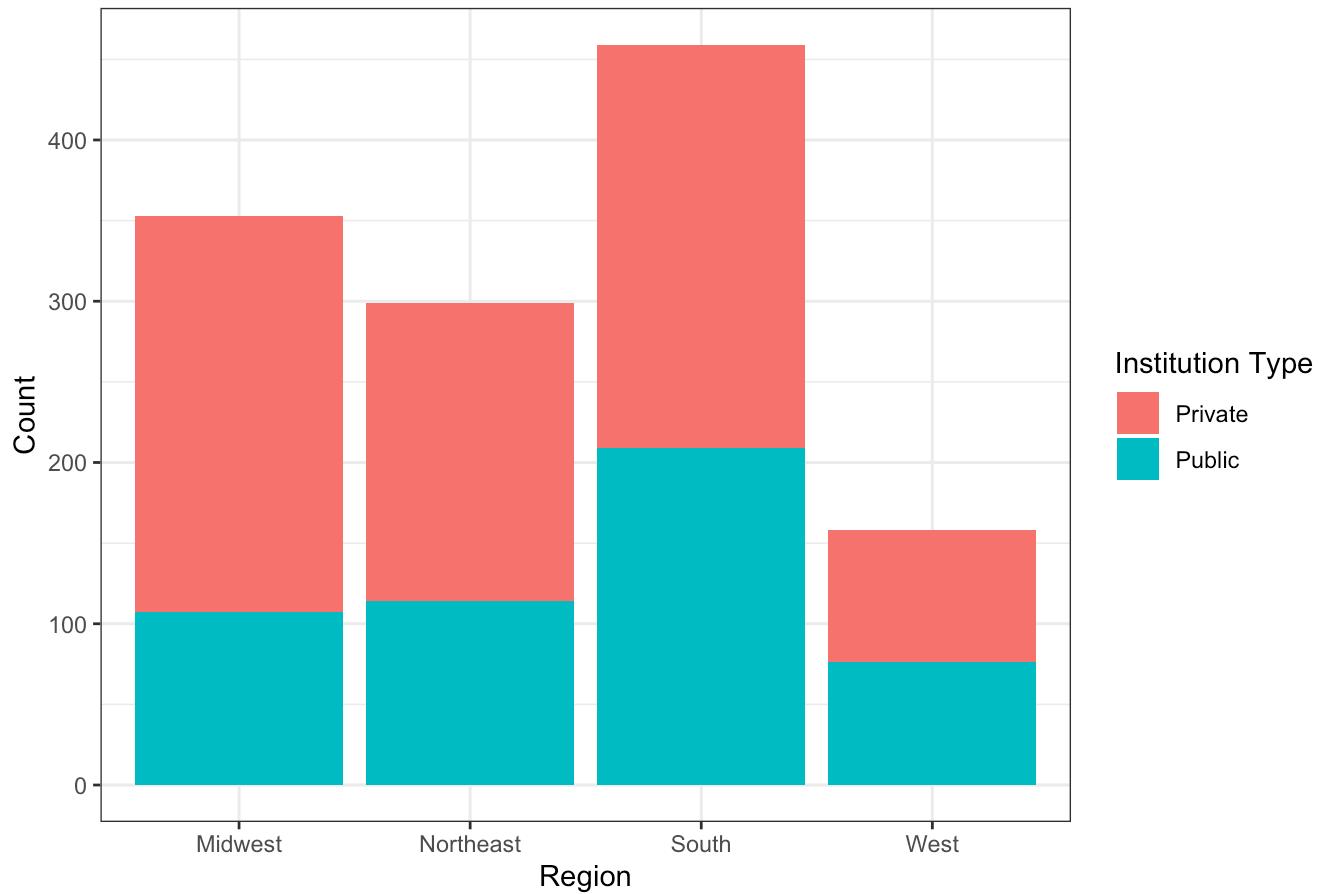
► Code

## College Type by Region



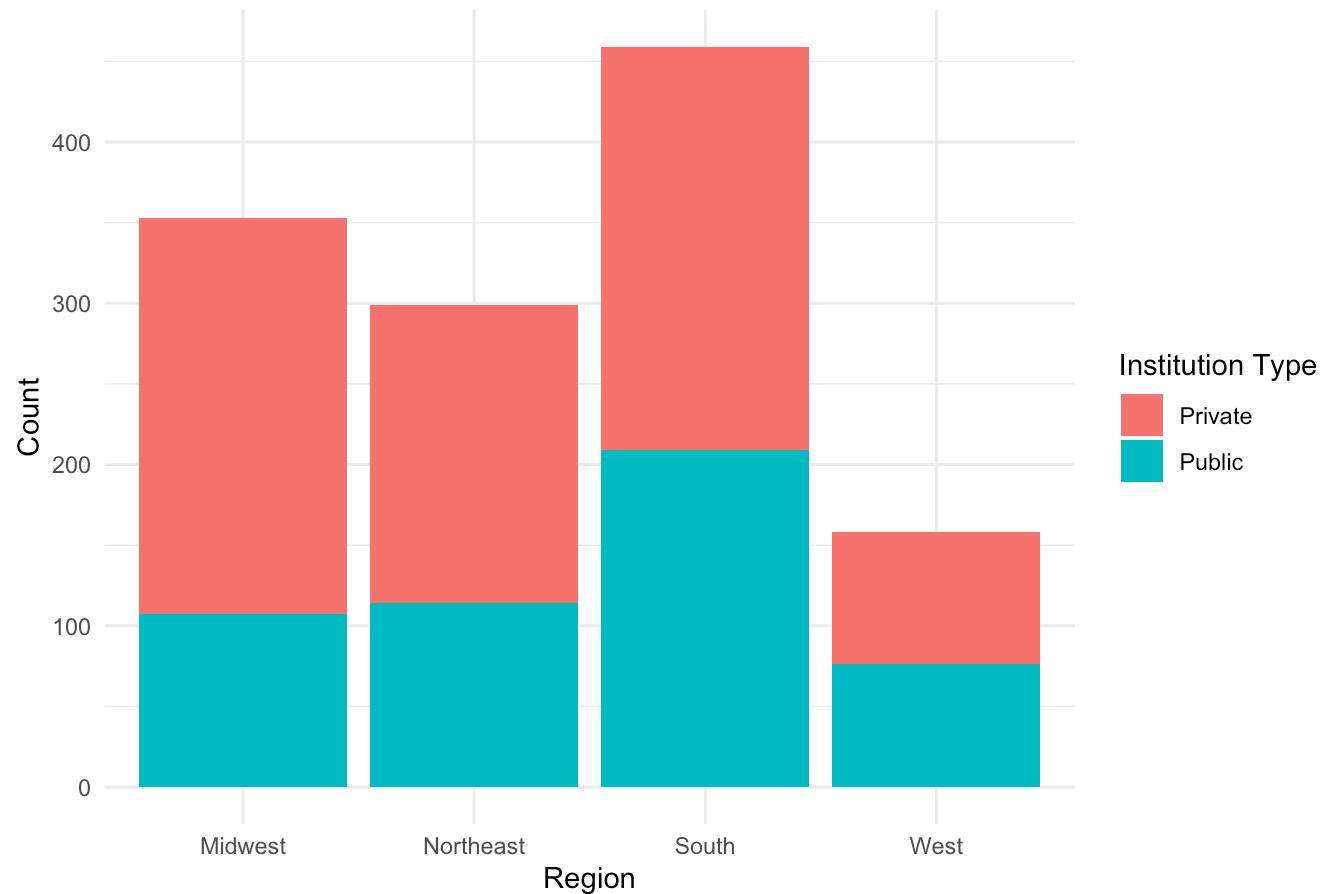
► Code

## College Type by Region



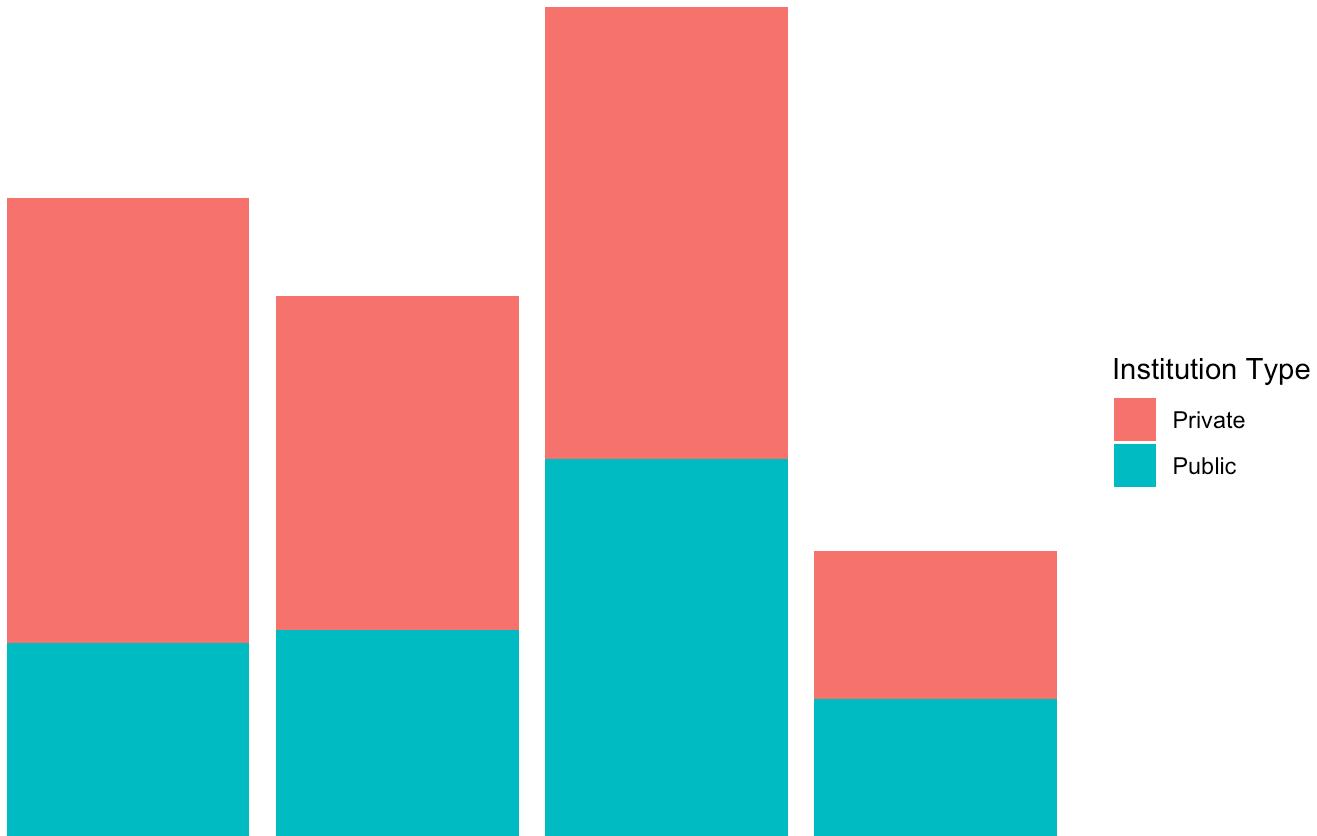
► [Code](#)

## College Type by Region



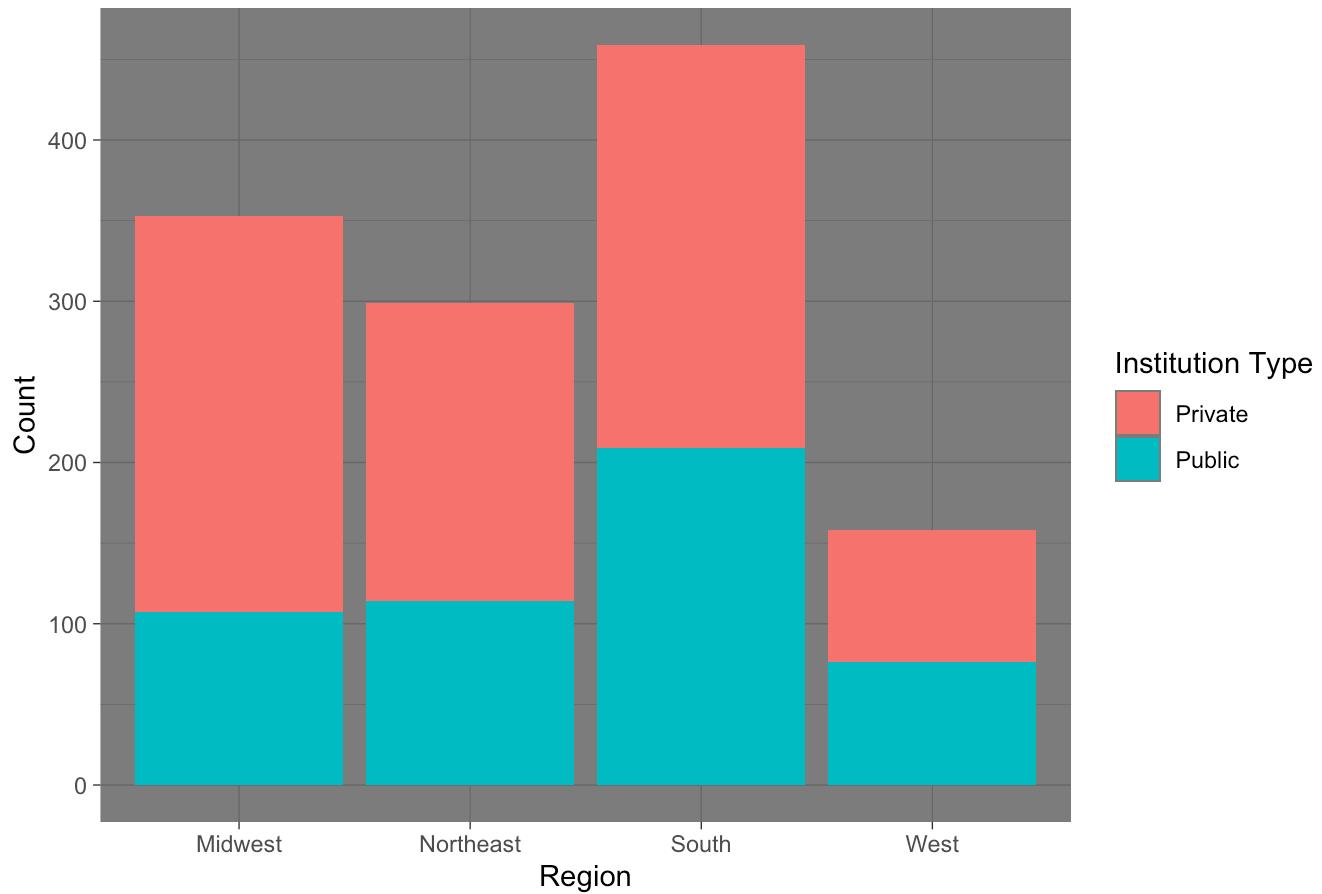
► Code

## College Type by Region



► Code

## College Type by Region



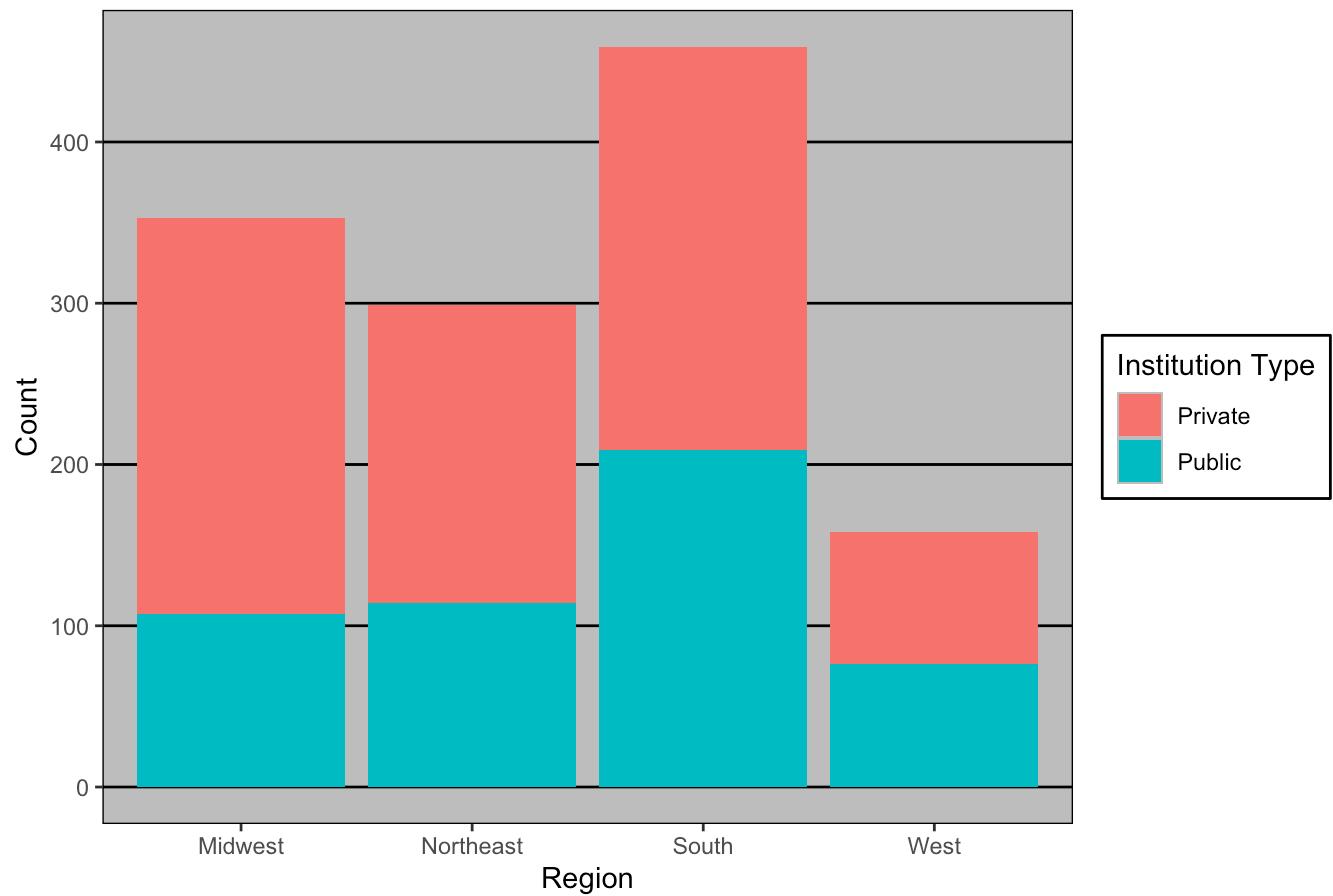
► Code

## College Type by Region



► [Code](#)

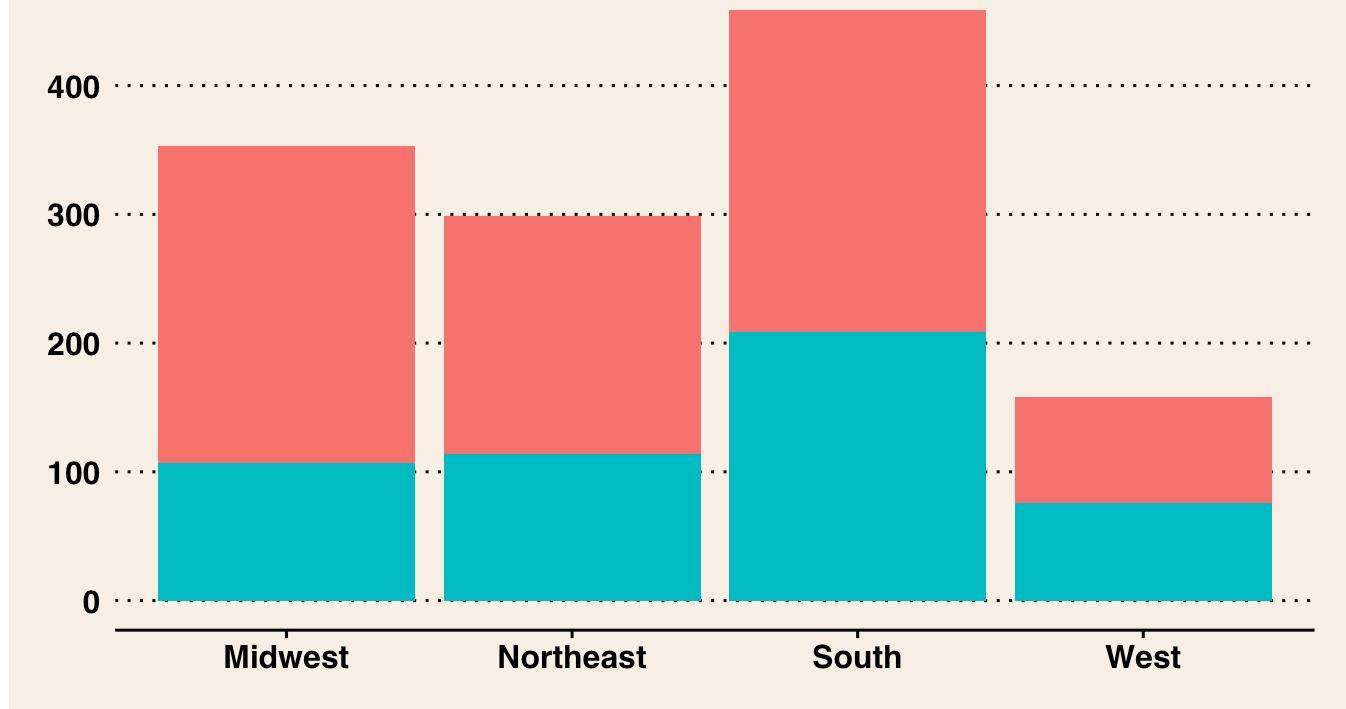
## College Type by Region



► Code

# College Type by Region

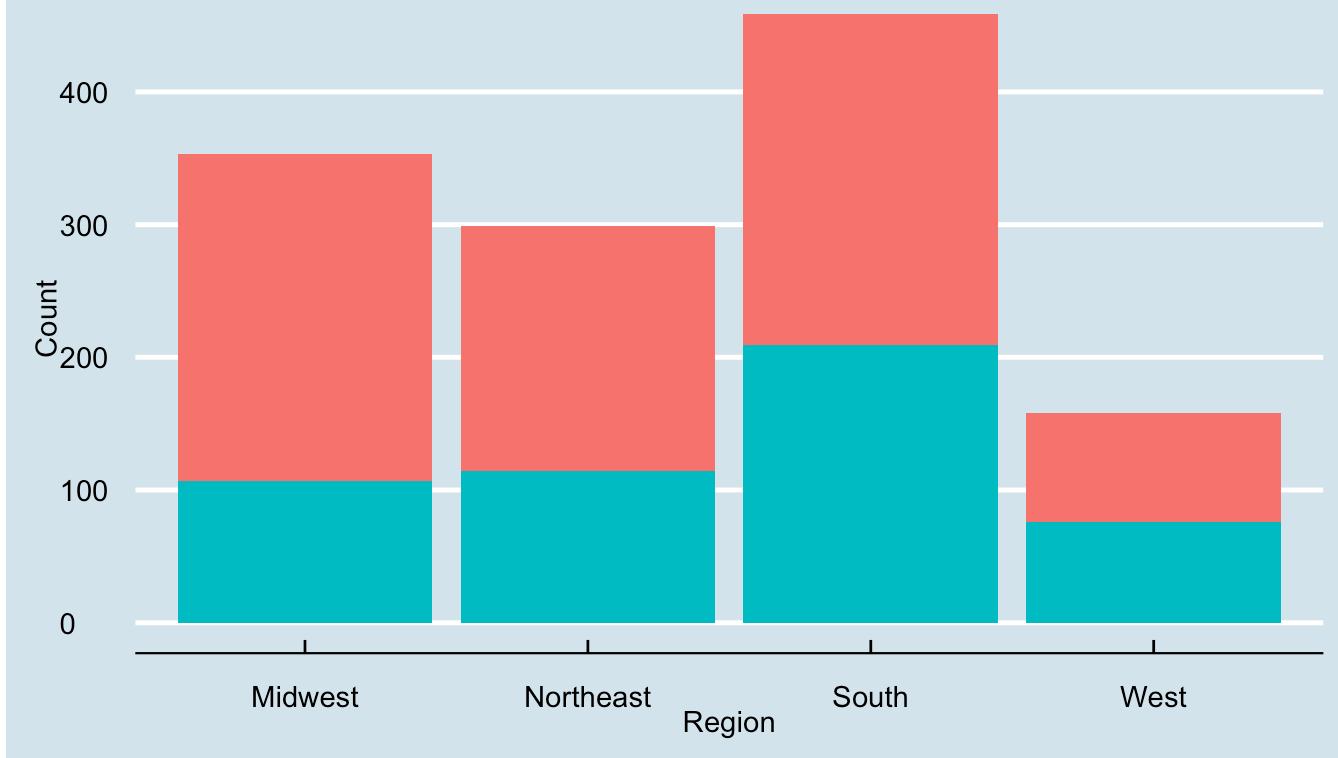
Institution Type ■ Private ■ Public



► Code

## College Type by Region

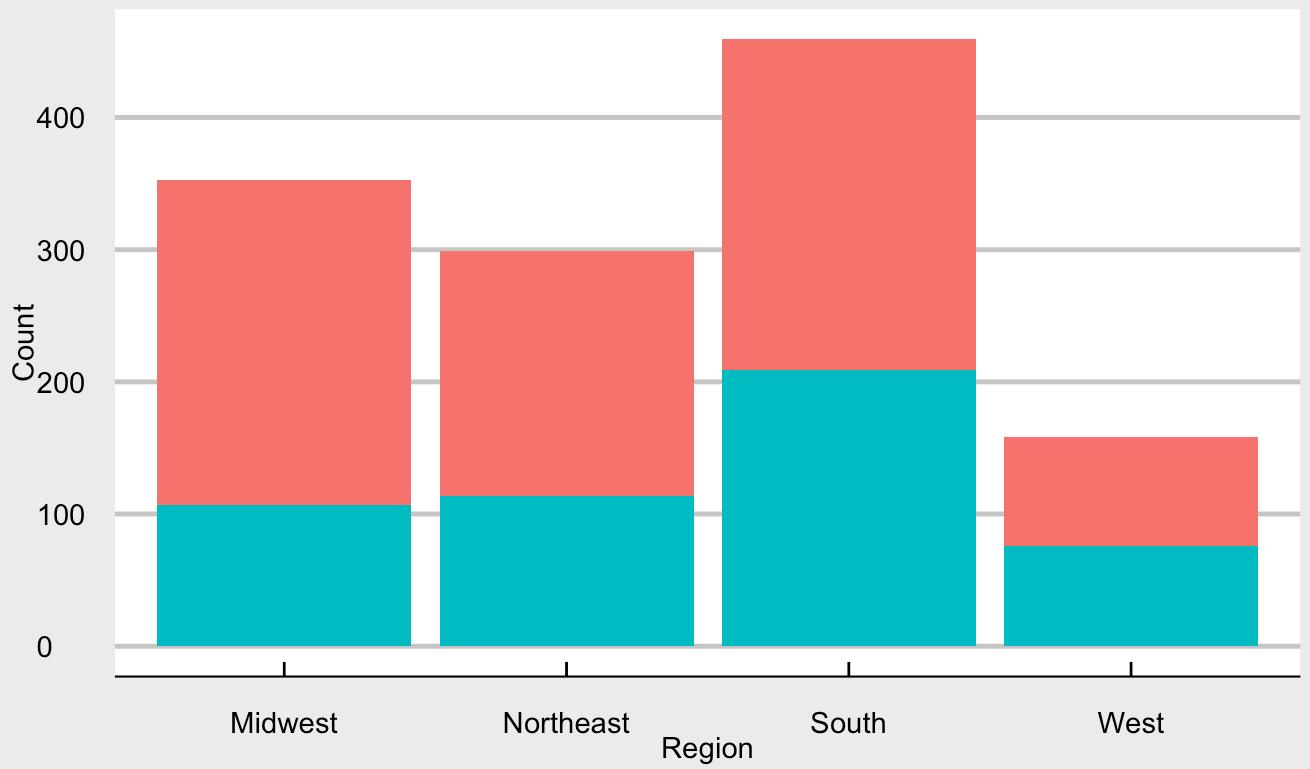
Institution Type    Private    Public



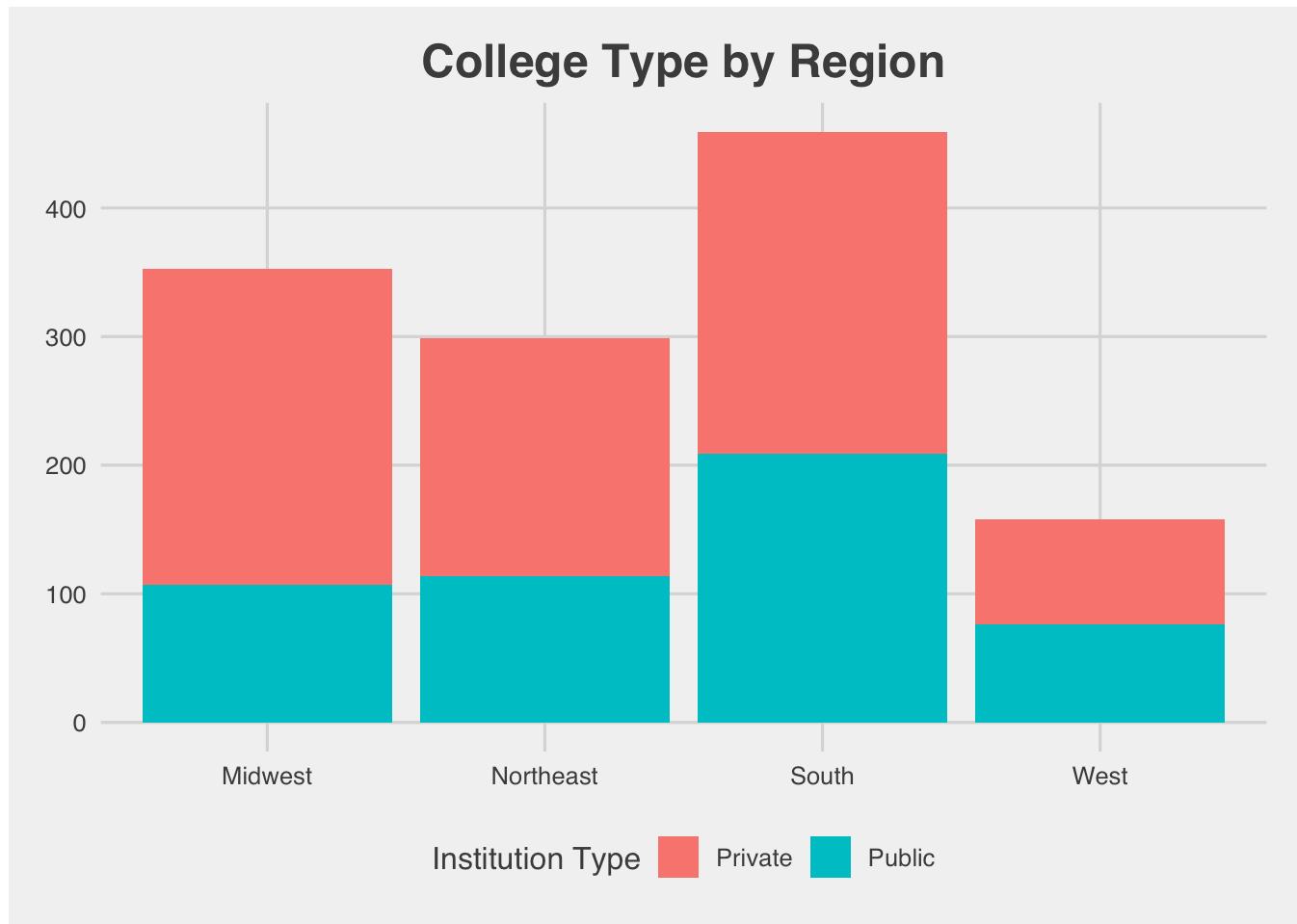
► Code

## College Type by Region

Institution Type    Private    Public



► Code



## 4 Part 3: Plotting Data Visualizations on Maps Using the Google API

► Code

## 5 Part 4: Interactive Data Visualizations

### 5.1 Deliverable 30: Prepare the Data We Will Plot

► Code

```
Rows: 2250 Columns: 8
-- Column specification --
Delimiter: ","
chr (4): Job Title, Category, Country, Region
dbl (4): Year, Total, Latitude, Longitude
```

- Use `spec()` to retrieve the full column specification for this data.
- Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

► Code

```
# A tibble: 10 × 8
  Year `Job Title` Category Country Region Total Latitude Longitude
  <dbl> <chr>     <chr>   <chr>   <chr>   <dbl>   <dbl>    <dbl>
1 2015 Software Engineer Engineering USA North ... 32 41.7 -84.2
2 2015 Data Scientist   Engineering USA North ... 33 29.8 -107.
3 2015 Marketing Manager Operations USA North ... 5 41.9 -111.
4 2015 Financial Analyst HR       USA North ... 48 43.0 -119.
5 2015 HR Specialist   Sales     USA North ... 28 25.6 -105.
6 2015 IT Support      Engineering USA North ... 71 40.9 -107.
7 2015 Sales Executive Marketing  USA North ... 34 39.4 -77.5
8 2015 Operations Manager Finance   USA North ... 17 40.8 -103.
9 2015 Legal Consultant  Finance   USA North ... 15 34.7 -85.6
10 2015 Medical Doctor   Legal     USA North ... 61 33.0 -102.
```

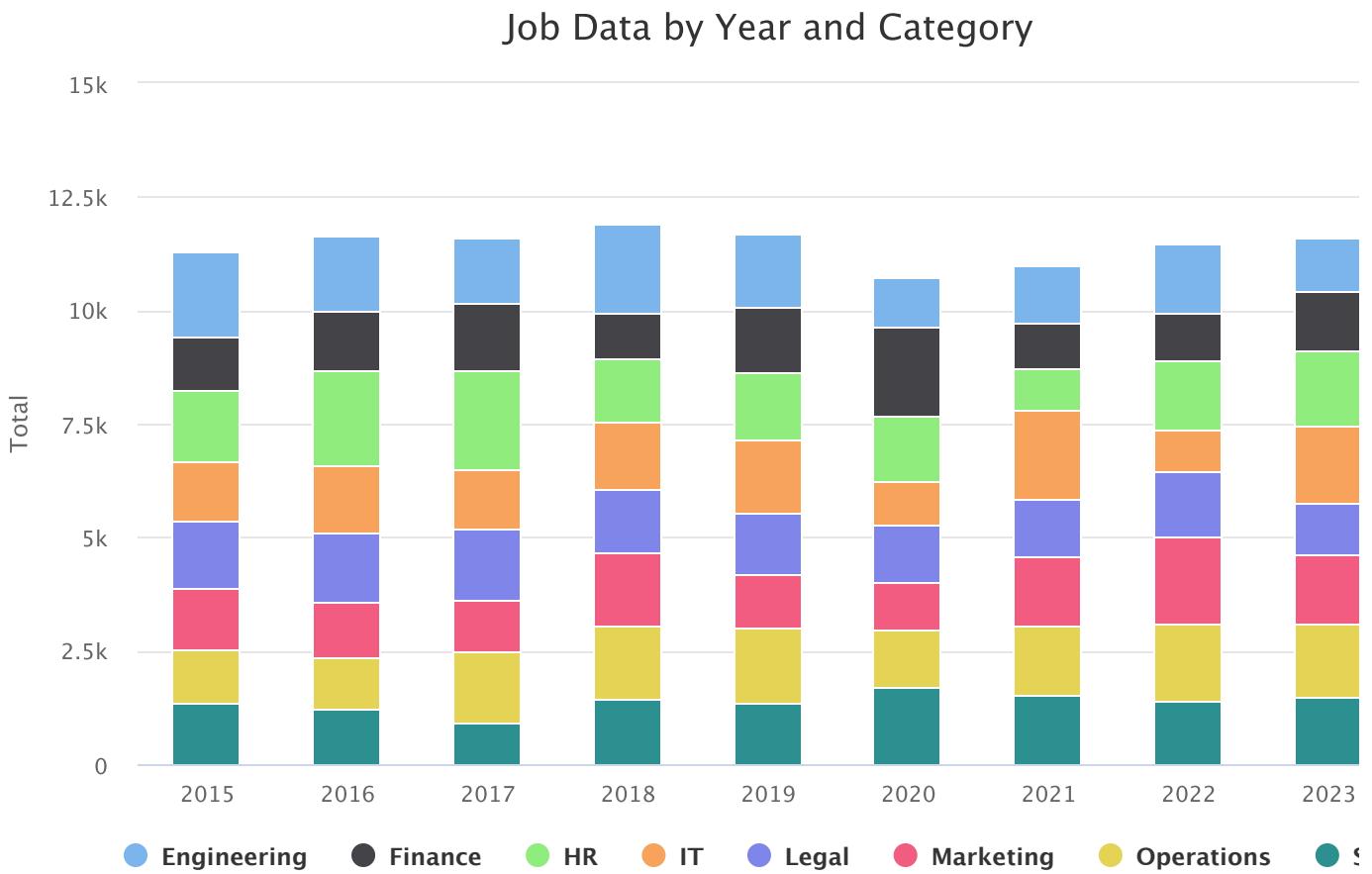
► Code

## 5.2 Deliverable 31: View Data in Highcharter

► Code

```
`summarise()` has grouped output by 'Year'. You can override using the
`.groups` argument.
```

► Code

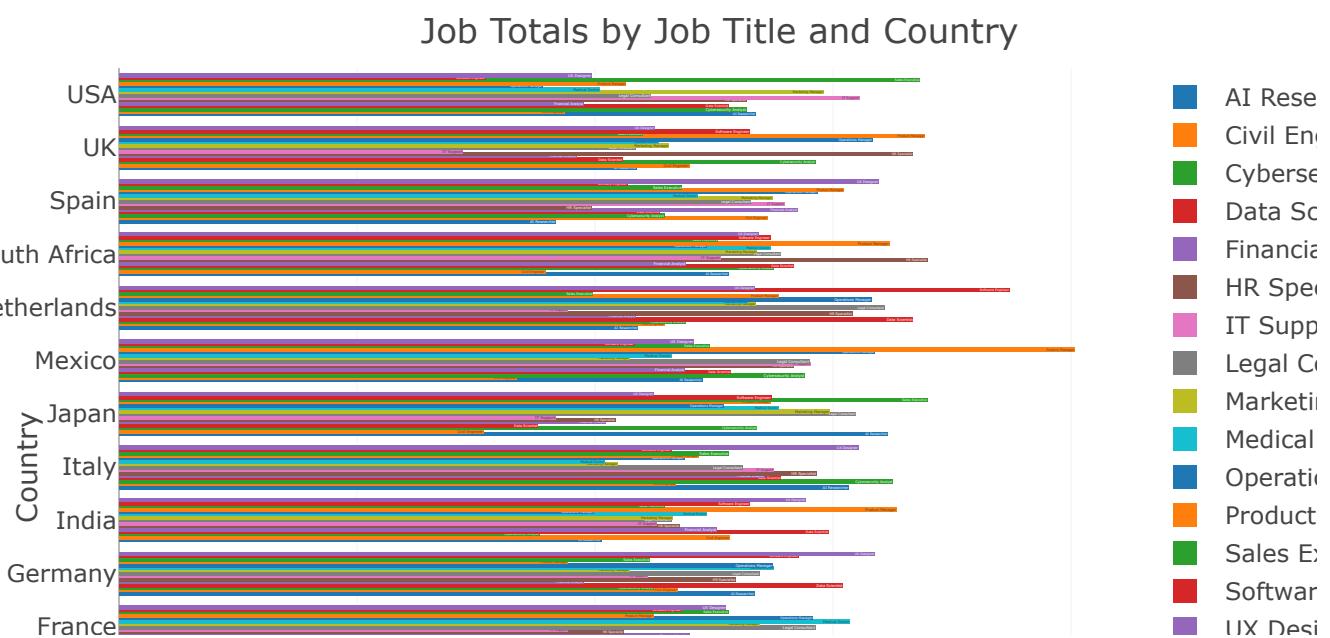


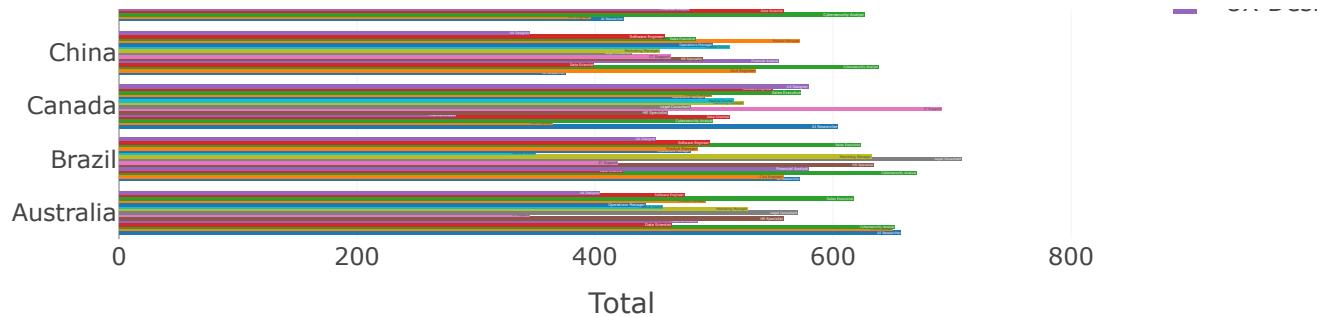
## 5.3 Deliverable 32: View Data in Plotly Stacked Bar Chart

### ► Code

```
`summarise()` has grouped output by 'Country'. You can override using the
`.groups` argument.
```

### ► Code



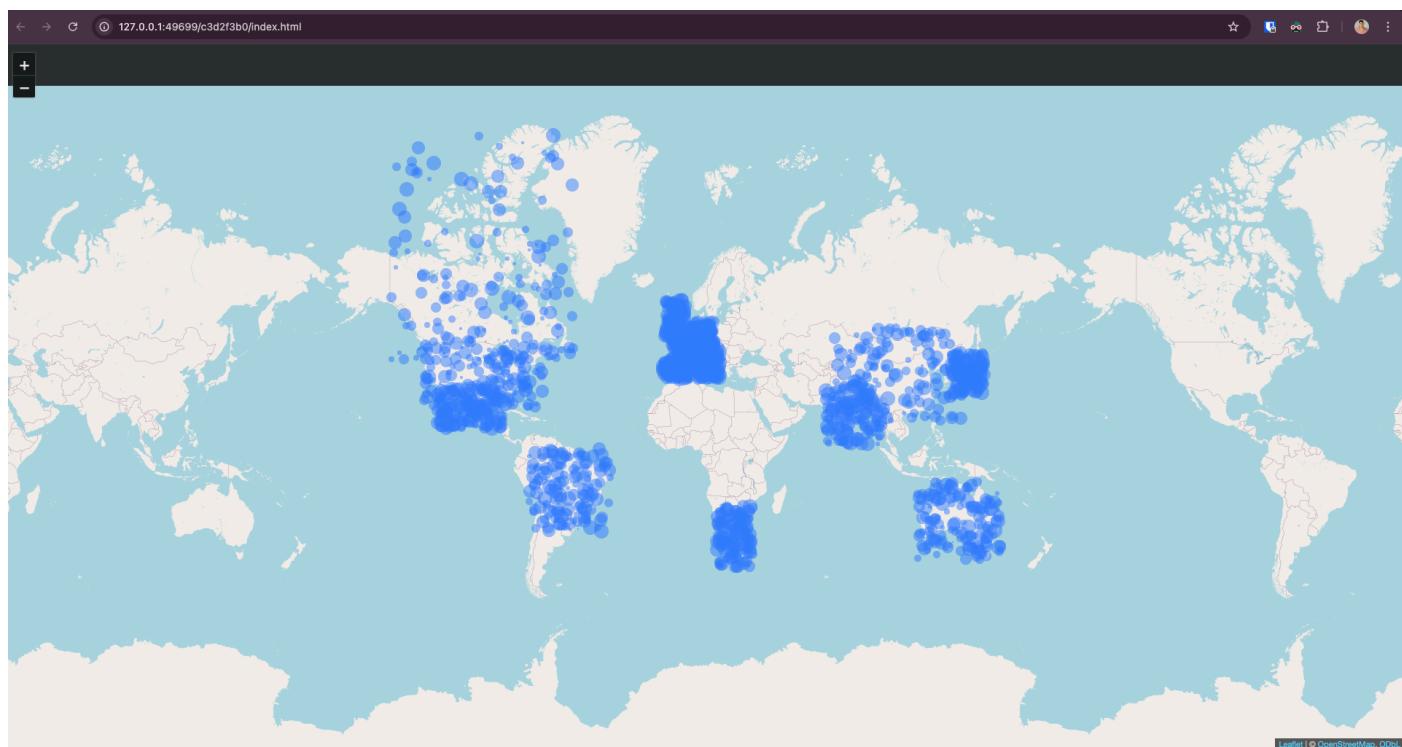


## 5.4 Deliverable 34: View Leaflet Map Plot

### 5.4.0.1 Note

I tried for about an hour to get this code to run, but it would not pick up the `Latitude` and `Longitude` columns in my data. I time capped this part of the lab at 1 hour and moved on. Screenshots of the maps are included below each code chunk.

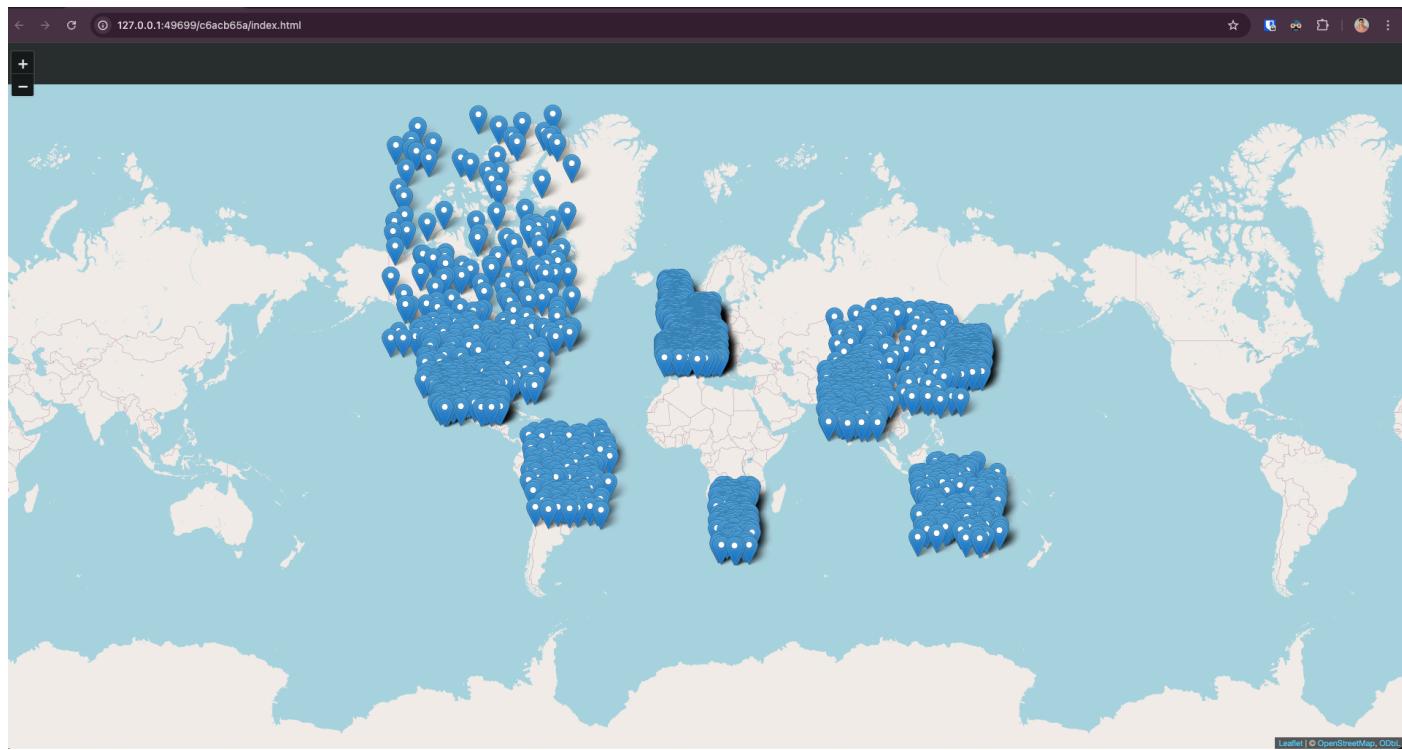
#### ► Code



Deliverable 34 Capture

## 5.5 Deliverable 35: View a Leaflet Map with Circle Markers

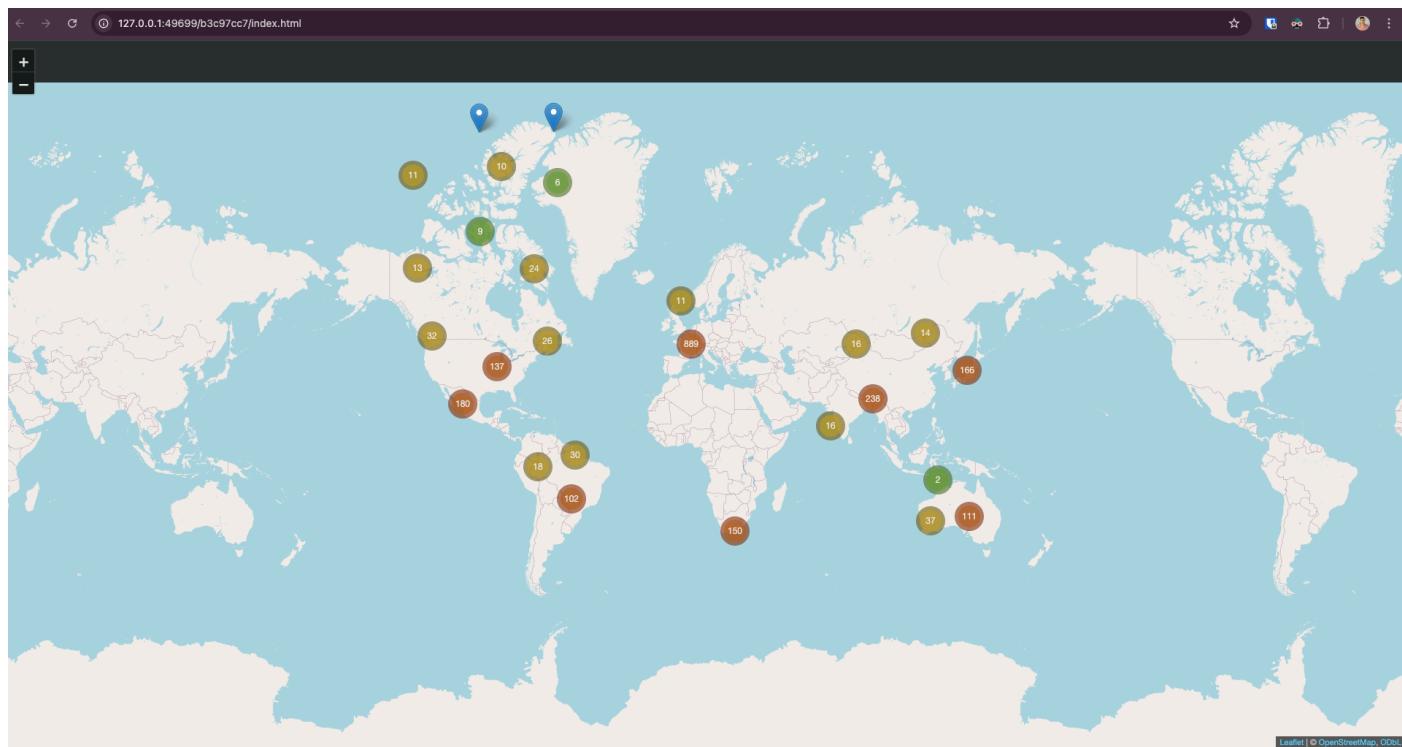
#### ► Code



Deliverable 35 Capture

## 5.6 Deliverable 36: View Leaflet Map with Clustered Markers

► Code



Deliverable 36 Capture

## 5.7 Deliverable 37: Read in Shapefiles and View in Chloropleth Map

► Code



Leaflet | © OpenStreetMap, ODbL

## 6 Part 5: Data Visualization in Python

### 6.1 Deliverable 38: Prepare Your Python Environment for DataViz

► Code

```
Hello, Python in Quarto
```

► Code

```
True
```

### 6.2 Deliverable 39: Prepare and Plot Your First Graph

► Code

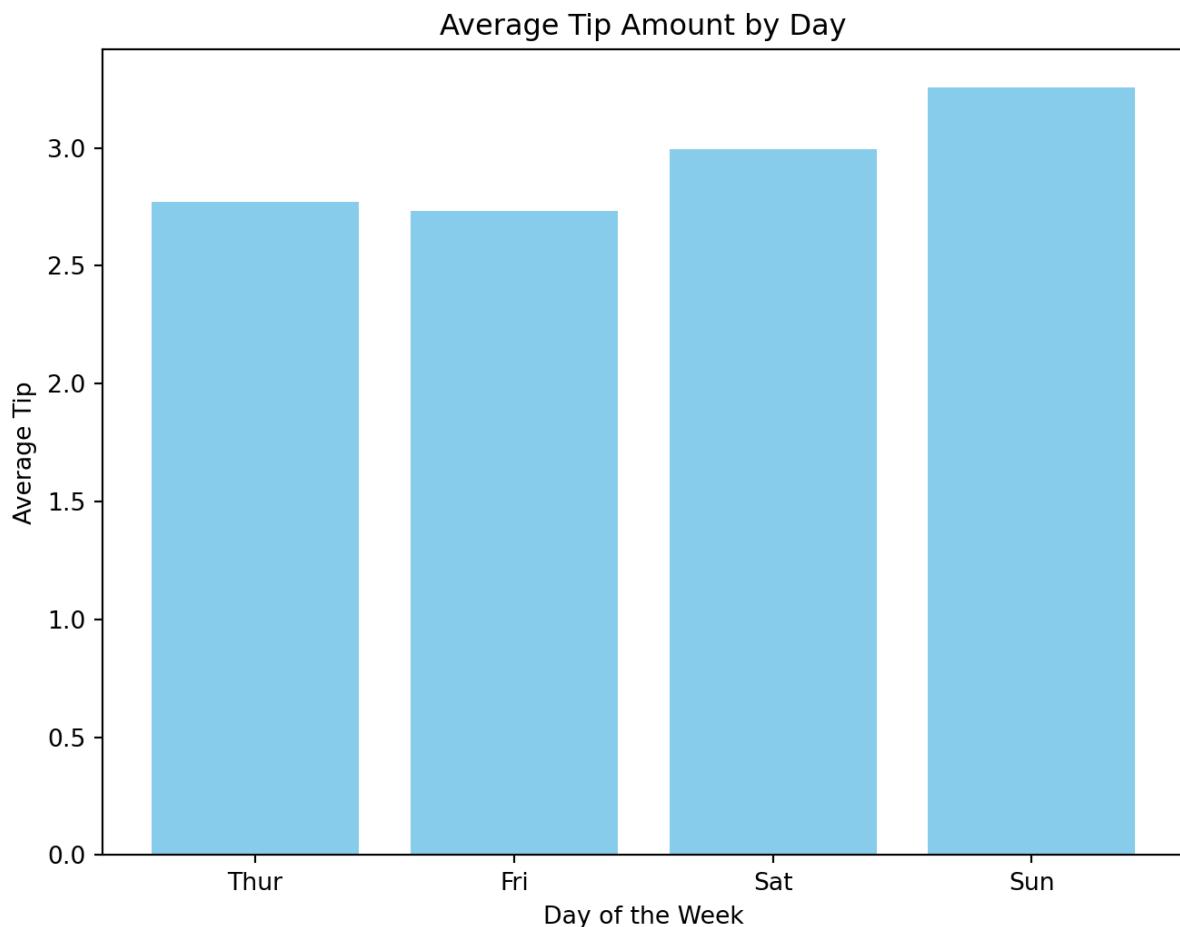
```
total_bill    tip      sex smoker  day     time   size
0         16.99  1.01  Female    No   Sun Dinner     2
1         10.34  1.66   Male     No   Sun Dinner     3
2         21.01  3.50   Male     No   Sun Dinner     3
```

```
3      23.68  3.31    Male     No   Sun Dinner      2
4      24.59  3.61  Female    No   Sun Dinner      4
```

► Code

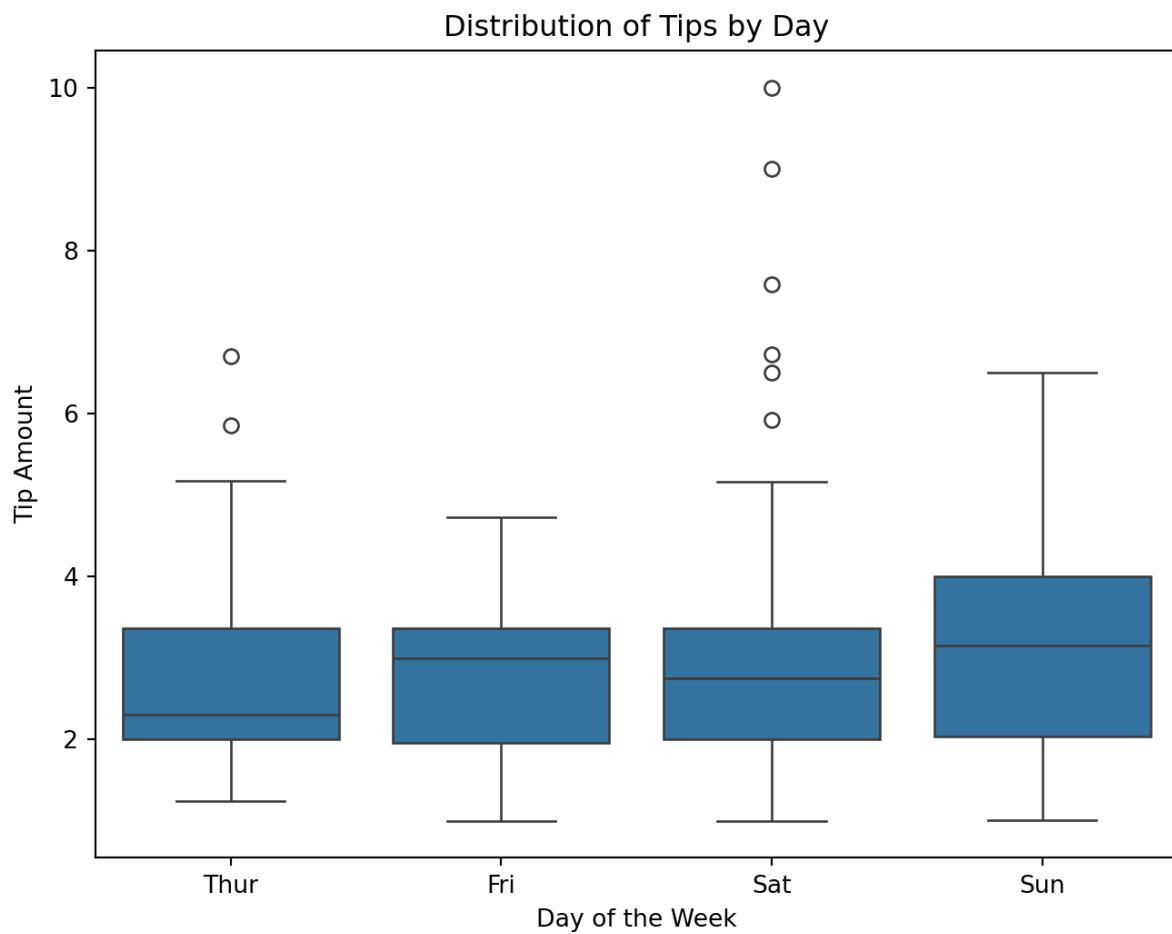
```
<string>:2: FutureWarning: The default of observed=False is deprecated and will be
changed to True in a future version of pandas. Pass observed=False to retain current
behavior or observed=True to adopt the future default and silence this warning.
```

► Code



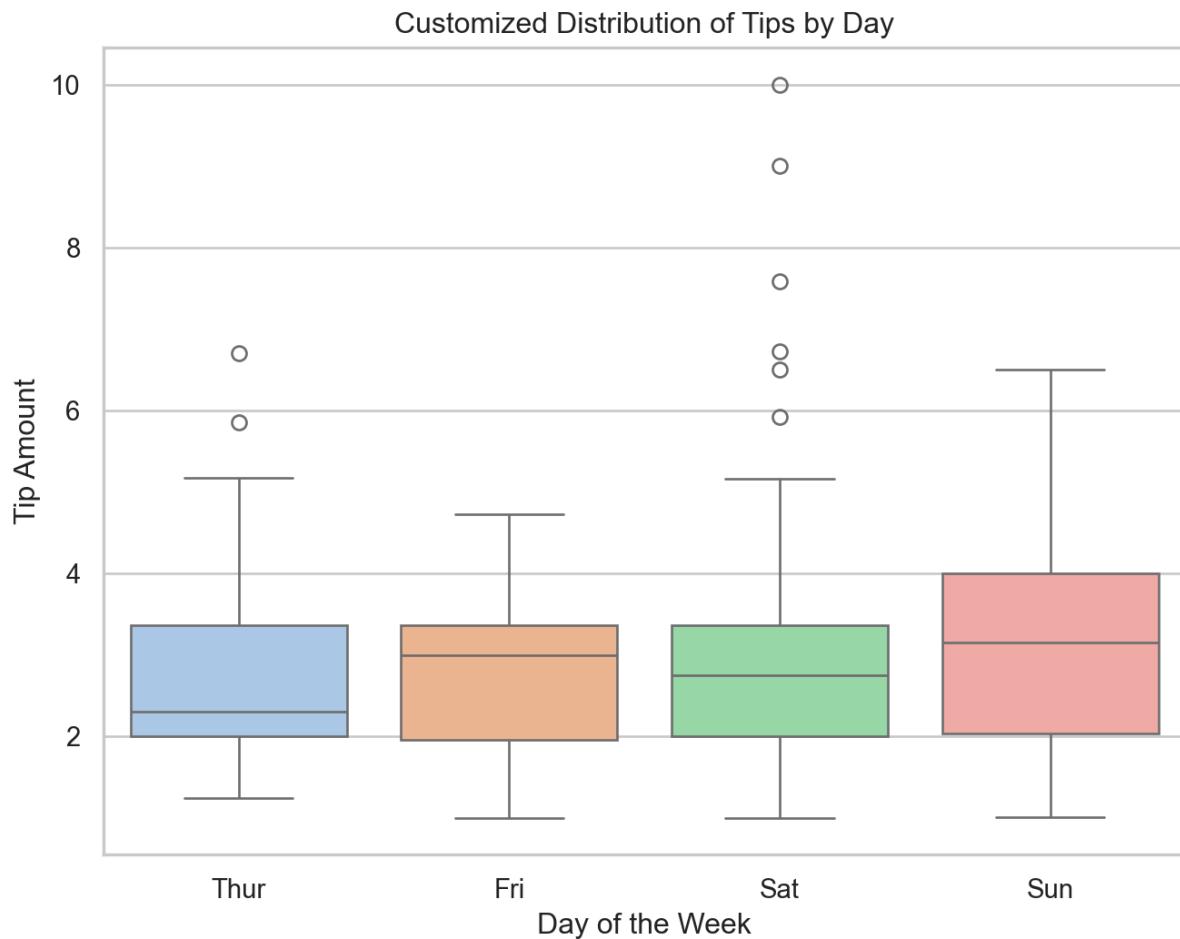
## 6.3 Deliverable 40: Data Visualization in Seaborn

► Code



## 6.4 Deliverable 41: Adjusting and Customizing your Data Visualization

► Code



## 6.5 Deliverable 42: Preparing to Visualize Text Data

### ► Code

My fellow citizens:

I stand here today humbled by the task before us, grateful for the trust you have bestowed, mindful of the sacrifices borne by our ancestors. I thank President Bush for his service to our nation, as well as the generosity and cooperation he has shown throughout this transition.

Forty-four Americans have now taken the presidential oath. The words have been spoken during rising tides of prosperity and the still waters of peace. Yet, every so often the oath is taken amidst gat

### ► Code

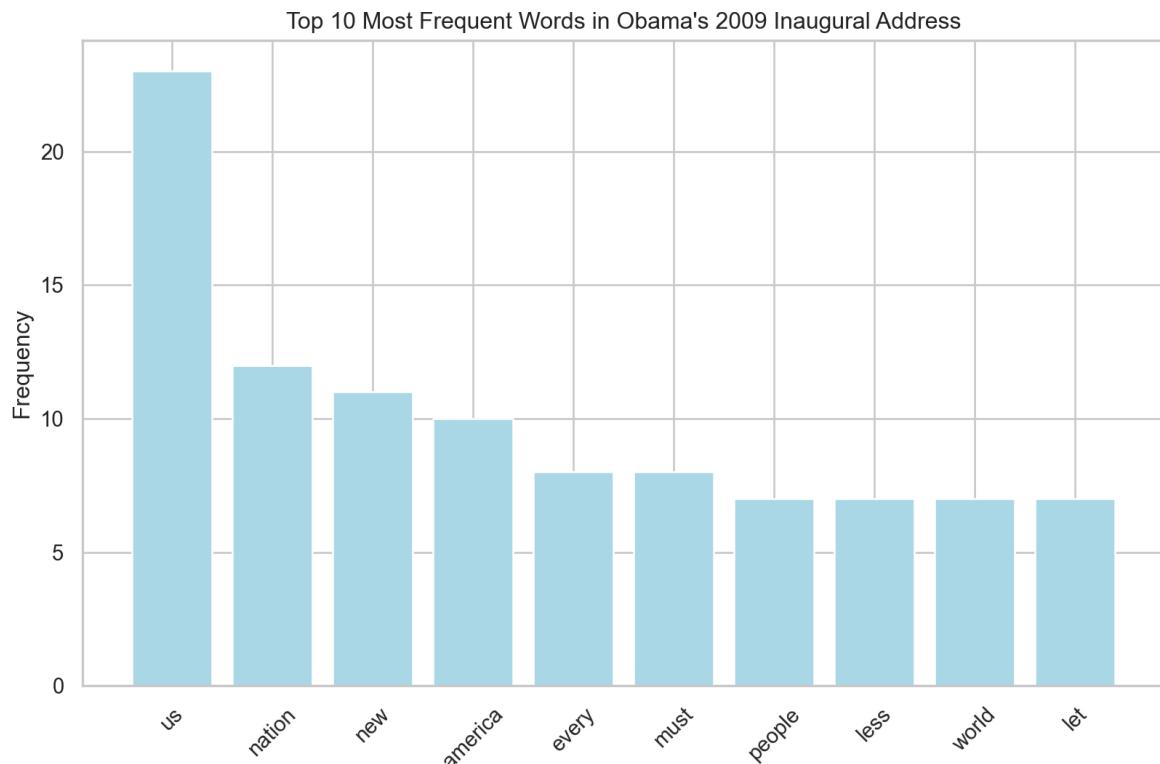
```
['fellow', 'citizens', 'stand', 'today', 'humbled', 'task', 'us', 'grateful', 'trust',
'bestowed', 'mindful', 'sacrifices', 'borne', 'ancestors', 'thank', 'president', 'bush',
'service', 'nation', 'well']
```

## 6.6 Deliverable 43: Visualizing Text Data

► Code

```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9], [Text(0, 0, 'us'), Text(1, 0, 'nation'), Text(2, 0, 'new'), Text(3, 0, 'america'), Text(4, 0, 'every'), Text(5, 0, 'must'), Text(6, 0, 'people'), Text(7, 0, 'less'), Text(8, 0, 'world'), Text(9, 0, 'let')])
```

► Code



## 7 Part 6: Building Dashboards in Shiny

### 7.1 Deliverable 44: Explore Shiny Examples

#### 7.1.0.1 Note

Ran a few of the shiny objects for practice, then commented them out for rendering.

► Code

```
Warning: package 'shiny' was built under R version 4.3.3
```

```
Attaching package: 'shiny'
```

The following objects are masked from 'package:DT':

dataTableOutput, renderDataTable

► Code

## 7.2 Deliverable 45: Creating a Basic RShiny App from Template

► Code

```
[1] "Ran shiny app through console with `shiny::runApp(\"Lab_7/app.R\")`"
```

## 7.3 Deliverable 46: Creating a Basic PyShiny App from Template

► Code

```
[1] "Ran shiny app through terminal with `shiny run --reload --port 8501 Lab_7/app.py`"
```