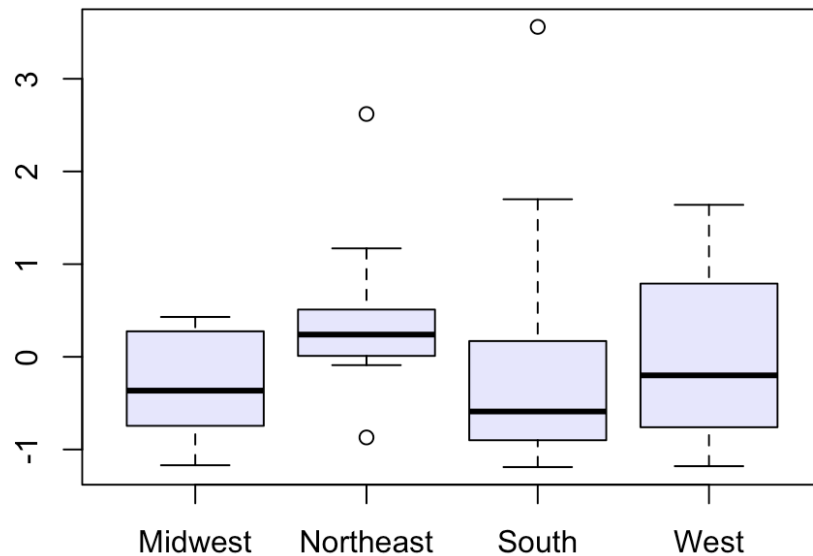


- **Bar charts**

First, we split data by region which creates new data frame

Name	Type	Value
viz.reg.dist	list [4]	List of length 4
Midwest	double [12]	0.43 0.37 -0.34 -0.87 -0.39 0.20 ...
Northeast	double [9]	0.32 -0.87 2.62 0.24 0.01 1.17 ...
South	double [17]	-0.69 -0.59 0.17 3.56 -0.80 0.07 ...
West	double [13]	-0.90 -0.20 1.24 0.79 -0.76 -0.38 ...

Then we draw boxplots by region

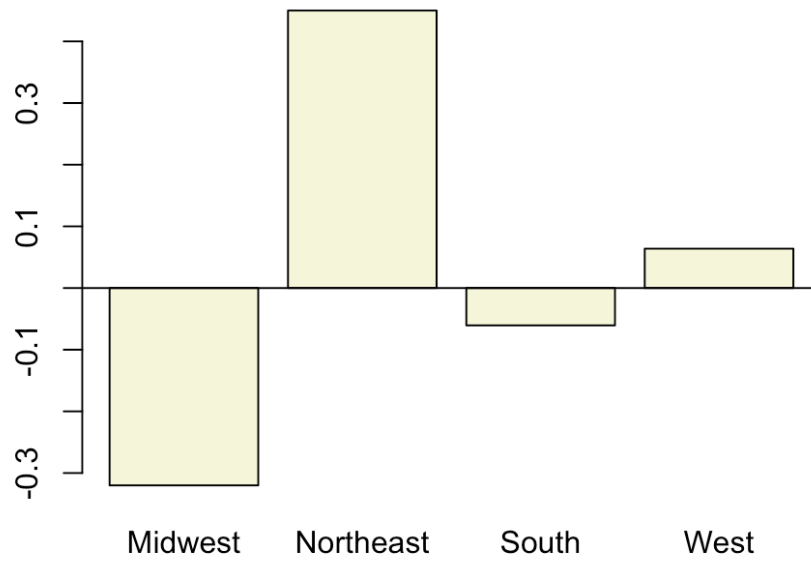


Following are means, which are included in the boxplot above

Values	
viz.reg.mean	Named num [1:4] -0.32 0.45 -0.0606 0.0638

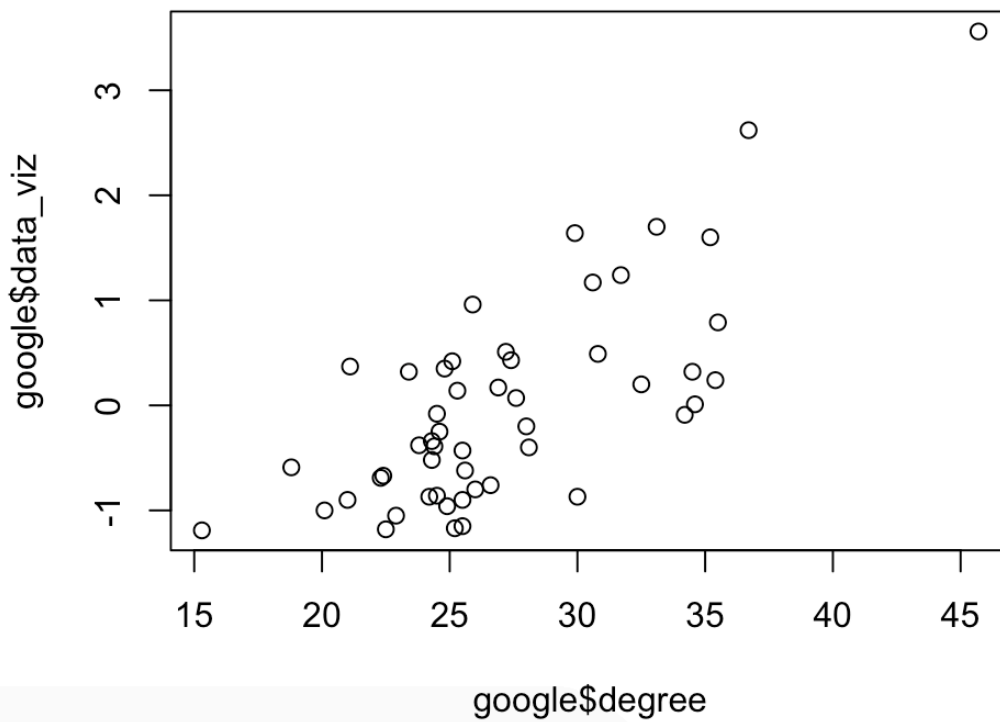
Next, we run them sequentially with adding the title

Average Google Search Share of "Data Visualization" by Region of US

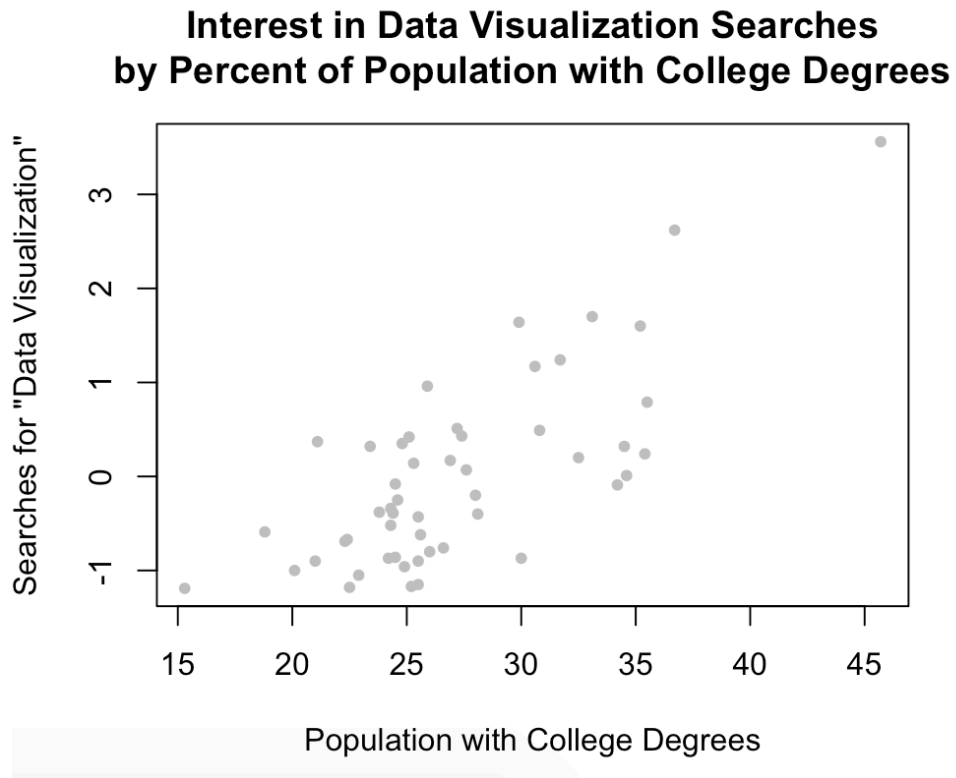


- **Scatterplots**

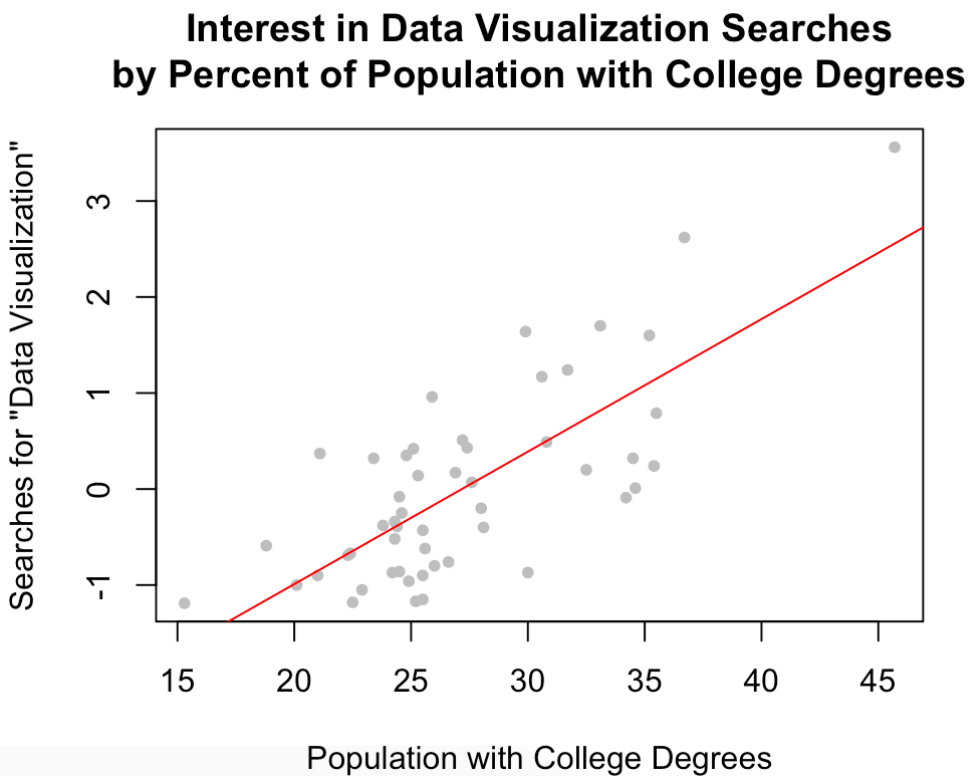
To check if there is an association between the percentage of people in a state with college degrees and the interest in data visualization, we created a scatterplot graph



Then we added title, labels and changed circles to points



Then we added a linear regression line



- **Calculating Correlations**

First we create data frame with only quantitative variables

Data	
g.quant	51 obs. of 4 variables
data_viz: num	-0.69 -0.9 -0.2 -0.59 1.24 0.79 0.32 0.17 3.56 ...
degree : num	22.3 25.5 28 18.8 31.7 35.5 34.5 26.9 45.7 26 ...
facebook: num	1.93 -0.52 -1.18 2.21 -1.28 -1.33 -0.14 -0.34 -...
nba : num	-0.9 -1.08 1.23 -1.1 1.65 -0.68 0.1 0.03 -0.06 0.29...

Then create correlation matrix for data frame

```
> cor(g.quant)
      data_viz    degree  facebook     nba
data_viz 1.0000000 0.7455229 -0.6335615 0.2317077
degree   0.7455229 1.0000000 -0.6440282 0.1303140
facebook -0.6335615 -0.6440282 1.0000000 -0.3652457
nba      0.2317077 0.1303140 -0.3652457 1.0000000
```

Then, to get correlation matrix and p-values we coerced it from data frame to matrix

```
> rcorr(as.matrix(g.quant))
      data_viz degree facebook  nba
data_viz   1.00   0.75   -0.63   0.23
degree     0.75   1.00   -0.64   0.13
facebook  -0.63  -0.64    1.00  -0.37
nba        0.23   0.13   -0.37   1.00

n= 51

P
      data_viz degree facebook  nba
data_viz          0.0000 0.0000 0.1018
degree  0.0000          0.0000 0.3621
facebook 0.0000 0.0000          0.0084
nba      0.1018 0.3621 0.0084
```

- **Computing regression**

To calculate regression,

```
> summary(reg1)
```

Call:

```
lm(formula = data_viz ~ degree + stats_ed + facebook + nba +  
    has_nba + region, data = google)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.14807	-0.42172	0.03191	0.41573	1.26782

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	-2.812431	0.767628	-3.664
degree	0.093233	0.027439	3.398
stats_edyes	0.341196	0.248521	1.373
facebook	-0.356782	0.166099	-2.148
nba	0.006228	0.119583	0.052
has_nbeyes	0.101530	0.245919	0.413
regionNortheast	0.321224	0.322915	0.995
regionSouth	0.286078	0.267237	1.071
regionWest	-0.021947	0.289024	-0.076

	Pr(> t)
(Intercept)	0.000691 ***
degree	0.001498 **
stats_edyes	0.177067
facebook	0.037528 *
nba	0.958712
has_nbeyes	0.681811
regionNortheast	0.325548
regionSouth	0.290506
regionWest	0.939832

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6447 on 42 degrees of freedom
Multiple R-squared: 0.651, Adjusted R-squared: 0.5846
F-statistic: 9.795 on 8 and 42 DF, p-value: 1.505e-07