DataAnalysisProject

Noah Medina

2023-01-13

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

## Including Plots

1. Load in graduation data

graduation <- read.csv('graduation.csv')

1. View the data

head(graduation,55)

## Community Narcotics\_Calls Completion\_Rate  
## 1 Allendale/Irvington/S. Hilton 28.7 73.9  
## 2 Baltimore City 59.1 78.1  
## 3 Beechfield/Ten Hills/West Hills 29.5 78.6  
## 4 Belair-Edison 46.1 81.6  
## 5 Brooklyn/Curtis Bay/Hawkins Point 127.8 69.3  
## 6 Cedonia/Frankford 24.9 77.5  
## 7 Cherry Hill 36.0 72.7  
## 8 Chinquapin Park/Belvedere 14.5 81.8  
## 9 Claremont/Armistead 12.0 80.2  
## 10 Clifton-Berea 76.4 70.8  
## 11 Cross-Country/Cheswolde 1.0 82.8  
## 12 Dickeyville/Franklintown 9.3 80.0  
## 13 Dorchester/Ashburton 39.0 78.7  
## 14 Downtown/Seton Hill 105.7 77.8  
## 15 Edmondson Village 73.5 75.6  
## 16 Fells Point 3.3 72.7  
## 17 Forest Park/Walbrook 23.1 80.8  
## 18 Glen-Fallstaff 96.2 77.8  
## 19 Greater Charles Village/Barclay 54.8 73.3  
## 20 Greater Govans 37.4 82.4  
## 21 Greater Mondawmin 113.6 71.7  
## 22 Greater Roland Park/Poplar Hill 3.7 91.7  
## 23 Greater Rosemont 113.5 78.2  
## 24 Greenmount East 121.2 81.3  
## 25 Hamilton 12.3 86.6  
## 26 Harbor East/Little Italy 50.3 71.4  
## 27 Harford/Echodale 14.0 79.7  
## 28 Highlandtown 24.8 75.0  
## 29 Howard Park/West Arlington 17.6 83.6  
## 30 Inner Harbor/Federal Hill 5.0 82.4  
## 31 Lauraville 28.2 83.0  
## 32 Loch Raven 15.4 77.5  
## 33 Madison/East End 186.8 66.7  
## 34 Medfield/Hampden/Woodberry/Remington 13.6 93.0  
## 35 Midtown 6.6 83.3  
## 36 Midway/Coldstream 70.6 75.0  
## 37 Morrell Park/Violetville 26.9 84.6  
## 38 Mount Washington/Coldspring 14.0 88.0  
## 39 North Baltimore/Guilford/Homeland 14.2 97.2  
## 40 Northwood 12.9 74.6  
## 41 Oldtown/Middle East 34.9 69.7  
## 42 Orangeville/East Highlandtown 43.3 85.4  
## 43 Patterson Park North and East 30.6 85.2  
## 44 Penn North/Reservoir Hill 162.8 75.0  
## 45 Pimlico/Arlington/Hilltop 132.1 81.1  
## 46 Poppleton/The Terraces/Hollins Market 71.5 66.7  
## 47 Sandtown-Winchester/Harlem Park 305.7 81.0  
## 48 South Baltimore 1.9 100.0  
## 49 Southeastern 79.4 83.3  
## 50 Southern Park Heights 44.1 72.3  
## 51 Southwest Baltimore 277.4 67.8  
## 52 The Waverlies 43.2 80.0  
## 53 Upton/Druid Heights 415.5 76.5  
## 54 Washington Village/Pigtown 111.0 80.0  
## 55 Westport/Mount Winans/Lakeland 35.4 76.5  
## Dirty\_Street Library\_card No\_Internet X.\_Children\_Poverty Female\_Led Vacant  
## 1 77.8 203.7 22.4 21.4 77.8 6.5  
## 2 60.6 251.1 17.6 27.8 51.0 7.7  
## 3 24.2 148.2 15.0 14.8 60.0 0.6  
## 4 74.8 326.1 16.3 31.8 70.2 3.3  
## 5 108.3 230.8 30.4 51.3 59.1 8.1  
## 6 18.7 202.9 13.7 23.9 50.0 1.2  
## 7 15.1 349.5 30.7 51.2 81.0 6.0  
## 8 23.4 227.9 7.5 17.2 54.1 0.8  
## 9 17.0 172.8 23.2 32.6 56.8 0.1  
## 10 193.7 322.3 25.8 33.8 50.8 23.7  
## 11 3.8 100.1 12.5 6.0 8.8 0.2  
## 12 7.1 142.2 26.8 17.7 73.1 1.9  
## 13 32.9 214.8 22.2 30.3 66.6 3.2  
## 14 50.2 309.3 5.1 22.4 54.1 9.3  
## 15 36.4 291.8 23.0 27.4 68.5 3.9  
## 16 26.9 240.1 5.4 4.3 14.1 0.9  
## 17 49.6 279.2 35.2 23.8 53.3 5.9  
## 18 27.0 185.8 22.6 26.4 36.7 1.1  
## 19 53.4 265.0 10.9 25.1 43.0 4.2  
## 20 45.9 304.3 15.9 41.3 51.7 3.6  
## 21 77.9 231.5 32.9 34.6 68.0 13.9  
## 22 8.1 415.1 3.4 0.0 8.1 0.1  
## 23 88.0 291.0 38.2 32.7 67.7 19.1  
## 24 161.0 295.6 28.6 44.1 69.6 29.2  
## 25 18.4 212.3 8.0 1.9 45.7 1.0  
## 26 54.9 318.1 11.5 34.8 71.4 1.6  
## 27 11.9 167.0 10.2 8.5 37.2 0.7  
## 28 99.3 291.2 10.6 27.0 35.9 1.2  
## 29 29.7 159.1 25.1 12.8 39.9 1.7  
## 30 18.0 301.2 7.9 4.2 11.2 0.4  
## 31 21.4 236.6 6.0 11.5 34.7 1.2  
## 32 25.7 185.8 14.9 13.1 56.8 0.4  
## 33 340.7 330.5 16.7 49.9 64.8 18.7  
## 34 25.9 311.8 9.1 6.7 20.8 0.7  
## 35 33.5 359.3 9.7 19.4 24.7 1.2  
## 36 154.6 306.1 27.0 33.1 66.7 19.8  
## 37 57.4 75.0 27.4 14.2 47.2 1.9  
## 38 24.7 246.8 8.9 2.5 17.2 0.2  
## 39 9.1 229.6 6.1 4.0 13.8 0.5  
## 40 23.5 254.5 15.3 22.0 39.1 0.5  
## 41 51.5 309.1 23.7 48.6 72.8 9.9  
## 42 112.3 202.8 15.2 8.2 31.9 1.1  
## 43 177.0 355.2 14.4 49.4 34.2 3.5  
## 44 96.4 306.4 22.0 29.5 48.9 15.8  
## 45 55.0 214.9 32.6 30.1 69.0 15.0  
## 46 92.2 237.5 25.1 63.6 84.0 9.3  
## 47 135.3 321.2 38.6 60.5 71.9 32.0  
## 48 12.9 256.8 3.5 NA 17.7 0.2  
## 49 46.8 147.8 16.4 26.3 33.1 1.1  
## 50 63.9 251.5 27.6 20.0 64.3 18.0  
## 51 228.1 223.1 27.6 55.5 58.9 29.9  
## 52 59.2 384.2 20.0 46.5 55.9 4.3  
## 53 79.7 299.1 22.0 69.8 86.6 27.0  
## 54 132.8 308.7 8.9 35.9 70.0 5.7  
## 55 64.3 140.1 21.7 51.3 73.1 7.4  
## Violent\_Crime Household\_Income num\_arts\_business Poverty Poverty\_Cat  
## 1 20.0 42714.6 NA 14.2 2  
## 2 16.1 52318.2 0.6 15.0 2  
## 3 9.8 55435.5 0.1 8.5 2  
## 4 15.9 50010.4 0.2 18.9 3  
## 5 17.6 32598.8 0.2 32.1 4  
## 6 13.3 51097.8 0.2 12.5 2  
## 7 19.1 24762.1 0.4 38.0 4  
## 8 9.0 59616.5 0.3 9.2 2  
## 9 16.9 40449.7 0.1 20.5 3  
## 10 30.9 39619.1 NA 17.2 3  
## 11 3.5 63371.3 0.5 4.7 1  
## 12 6.5 45807.8 0.8 6.1 2  
## 13 14.2 55345.8 NA 12.3 2  
## 14 50.0 59158.5 4.1 7.5 2  
## 15 14.1 44911.3 NA 14.5 2  
## 16 10.2 97869.6 1.6 3.3 1  
## 17 12.2 48992.1 0.2 15.9 3  
## 18 12.6 44228.1 0.6 14.3 2  
## 19 16.3 45797.7 1.6 12.2 2  
## 20 13.7 43781.9 0.1 17.2 3  
## 21 23.4 40488.6 0.3 10.5 2  
## 22 4.1 125868.9 1.1 0.2 1  
## 23 26.7 37490.0 0.1 20.9 3  
## 24 27.7 32207.9 0.3 25.8 3  
## 25 8.6 72534.9 0.5 4.5 1  
## 26 24.7 61235.2 1.6 24.8 3  
## 27 9.4 61434.2 0.4 4.8 1  
## 28 12.7 99549.1 0.7 7.8 2  
## 29 10.2 50991.6 0.2 9.7 2  
## 30 7.4 104058.3 1.3 2.9 1  
## 31 9.0 70883.2 0.5 7.6 2  
## 32 9.2 55191.4 0.1 8.3 2  
## 33 36.1 37302.9 0.2 32.1 4  
## 34 8.7 72756.9 1.4 3.7 1  
## 35 13.5 58809.8 3.3 6.1 2  
## 36 29.4 35897.9 0.1 24.2 3  
## 37 21.5 43842.8 0.6 8.2 2  
## 38 3.1 86714.8 0.7 2.9 1  
## 39 2.6 114246.7 0.6 5.1 2  
## 40 11.1 56756.6 0.5 7.3 2  
## 41 32.8 24190.7 0.7 41.4 4  
## 42 22.2 69139.1 0.5 8.7 2  
## 43 12.8 80328.0 0.2 18.0 3  
## 44 18.6 33920.8 0.5 11.4 2  
## 45 25.7 34041.7 0.1 18.9 3  
## 46 20.5 23373.6 0.6 45.8 4  
## 47 33.1 26689.5 0.1 30.2 4  
## 48 2.1 124827.1 0.7 0.0 1  
## 49 17.3 41980.9 0.6 23.2 3  
## 50 19.1 30312.4 0.1 29.0 3  
## 51 37.8 29768.3 0.4 34.3 4  
## 52 12.7 43890.4 1.0 25.3 3  
## 53 24.2 21530.5 0.4 43.5 4  
## 54 25.4 48941.5 1.3 24.4 3  
## 55 18.6 35200.6 0.1 28.5 3  
## Rent\_Affordability Not\_in\_Labor\_Force Percent\_Black Percent\_White  
## 1 47.5 28.6 85.2 6.9  
## 2 47.5 28.8 57.3 26.9  
## 3 56.8 20.0 74.9 15.7  
## 4 63.1 31.7 86.2 7.8  
## 5 52.1 35.4 35.8 31.7  
## 6 46.7 25.7 82.6 9.7  
## 7 40.8 37.5 89.1 2.5  
## 8 41.0 18.4 67.4 22.4  
## 9 42.8 29.8 53.2 18.8  
## 10 51.5 33.1 89.9 2.3  
## 11 41.7 22.2 22.9 66.8  
## 12 43.3 24.1 79.3 11.3  
## 13 47.9 26.4 90.9 2.3  
## 14 50.6 25.3 37.8 34.2  
## 15 67.5 25.6 93.7 1.3  
## 16 35.3 12.2 6.6 67.4  
## 17 53.8 28.4 89.8 3.6  
## 18 53.2 26.1 57.0 25.7  
## 19 49.2 38.9 30.3 38.6  
## 20 59.0 25.9 86.5 6.9  
## 21 52.6 34.9 87.5 3.8  
## 22 50.0 21.2 8.8 73.9  
## 23 50.5 39.0 94.2 1.2  
## 24 54.2 40.6 93.2 1.5  
## 25 31.1 25.0 62.2 28.1  
## 26 30.6 28.6 44.1 34.1  
## 27 40.3 18.1 58.8 30.4  
## 28 33.0 9.5 6.4 59.9  
## 29 71.3 33.3 89.5 2.4  
## 30 38.4 14.5 10.8 73.8  
## 31 34.7 26.8 59.3 29.8  
## 32 42.6 21.4 87.0 5.5  
## 33 55.4 43.6 77.5 4.8  
## 34 39.5 13.1 9.5 71.8  
## 35 38.5 18.9 29.9 46.8  
## 36 54.1 36.6 90.5 2.5  
## 37 47.1 34.5 28.4 51.0  
## 38 45.2 24.2 23.0 63.9  
## 39 47.8 39.0 12.1 59.2  
## 40 55.2 37.7 78.3 10.2  
## 41 48.3 52.9 77.5 8.7  
## 42 29.7 26.2 10.1 32.4  
## 43 46.0 18.6 27.8 39.1  
## 44 60.7 32.6 82.2 9.6  
## 45 49.7 28.8 90.7 3.1  
## 46 42.8 34.8 78.4 10.6  
## 47 53.1 42.2 92.5 2.0  
## 48 25.8 11.1 4.6 81.5  
## 49 64.9 35.7 24.5 28.9  
## 50 45.9 38.8 90.5 3.7  
## 51 62.3 41.2 75.2 12.3  
## 52 57.0 27.4 68.5 19.1  
## 53 58.3 40.5 88.0 5.0  
## 54 45.7 20.7 51.8 29.8  
## 55 54.6 32.4 51.6 10.7

1. View a summary of the data

summary(graduation)

## Community Narcotics\_Calls Completion\_Rate Dirty\_Street   
## Length:55 Min. : 1.00 Min. : 66.70 Min. : 3.80   
## Class :character 1st Qu.: 14.35 1st Qu.: 74.80 1st Qu.: 23.85   
## Mode :character Median : 36.00 Median : 78.70 Median : 50.20   
## Mean : 64.59 Mean : 79.12 Mean : 66.14   
## 3rd Qu.: 77.90 3rd Qu.: 82.60 3rd Qu.: 83.85   
## Max. :415.50 Max. :100.00 Max. :340.70   
##   
## Library\_card No\_Internet X.\_Children\_Poverty Female\_Led   
## Min. : 75.0 Min. : 3.40 Min. : 0.00 Min. : 8.10   
## 1st Qu.:208.0 1st Qu.:10.40 1st Qu.:14.35 1st Qu.:36.30   
## Median :251.5 Median :16.70 Median :26.70 Median :54.10   
## Mean :253.5 Mean :18.53 Mean :27.83 Mean :50.79   
## 3rd Qu.:307.6 3rd Qu.:25.45 3rd Qu.:35.62 3rd Qu.:68.25   
## Max. :415.1 Max. :38.60 Max. :69.80 Max. :86.60   
## NA's :1   
## Vacant Violent\_Crime Household\_Income num\_arts\_business  
## Min. : 0.100 Min. : 2.10 Min. : 21531 Min. :0.1000   
## 1st Qu.: 0.950 1st Qu.:10.00 1st Qu.: 37396 1st Qu.:0.2000   
## Median : 3.300 Median :15.90 Median : 48942 Median :0.5000   
## Mean : 7.025 Mean :17.16 Mean : 54260 Mean :0.6549   
## 3rd Qu.: 9.300 3rd Qu.:22.80 3rd Qu.: 61335 3rd Qu.:0.7000   
## Max. :32.000 Max. :50.00 Max. :125869 Max. :4.1000   
## NA's :4   
## Poverty Poverty\_Cat Rent\_Affordability Not\_in\_Labor\_Force  
## Min. : 0.00 Min. :1.000 Min. :25.80 Min. : 9.50   
## 1st Qu.: 7.55 1st Qu.:2.000 1st Qu.:42.15 1st Qu.:23.15   
## Median :14.20 Median :2.000 Median :47.90 Median :28.60   
## Mean :16.26 Mean :2.418 Mean :48.15 Mean :28.88   
## 3rd Qu.:24.30 3rd Qu.:3.000 3rd Qu.:54.15 3rd Qu.:35.55   
## Max. :45.80 Max. :4.000 Max. :71.30 Max. :52.90   
##   
## Percent\_Black Percent\_White   
## Min. : 4.60 Min. : 1.20   
## 1st Qu.:30.10 1st Qu.: 4.90   
## Median :68.50 Median :15.70   
## Mean :59.66 Mean :24.62   
## 3rd Qu.:87.25 3rd Qu.:34.15   
## Max. :94.20 Max. :81.50   
##

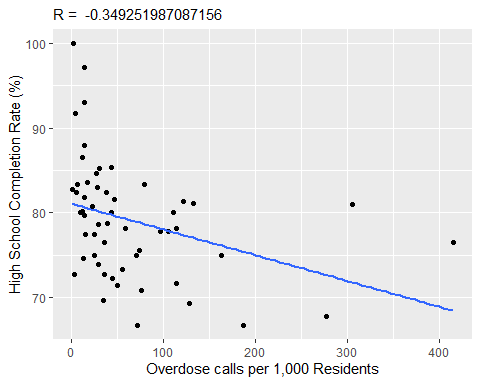
1. R calculation

calculate\_R <- function(column) {  
 c <- column  
 n <- 55  
 x <- 0  
 y <- 0  
 xx <- 0  
 yy <- 0  
 xy <- 0  
 for (i in 1:n) {  
 x = x + graduation[i,c]  
 y <- y + graduation[i,3]  
 xx <- xx + graduation[i,c]^2  
 yy <- yy + graduation[i,3]^2  
 xy <- xy + graduation[i,c] \* graduation[i,3]  
 }  
 R <- (n\*xy - x\*y) / sqrt((n\*xx - x^2) \* (n\*yy - y^2))  
 return(R)  
}

1. Overdose Calls vs Completion Rate

ggplot(data = graduation, aes(x = Narcotics\_Calls, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +  
 labs(x = "Overdose calls per 1,000 Residents", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(2)))

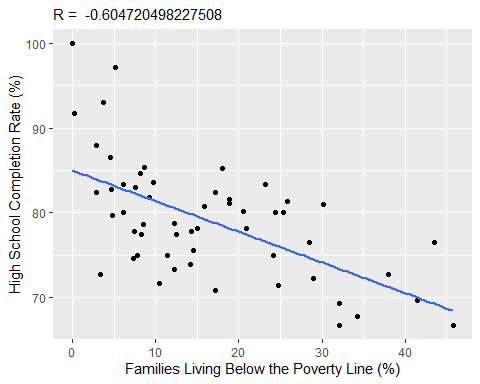
## `geom\_smooth()` using formula = 'y ~ x'



1. Families below the poverty line vs Completion Rate

ggplot(data = graduation, aes(x = Poverty, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +  
 labs(x = "Families Living Below the Poverty Line (%)", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(13)))

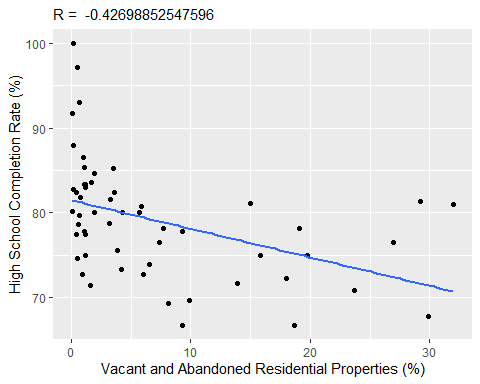
## `geom\_smooth()` using formula = 'y ~ x'



1. Vacant and abandoned residential properties vs Completion Rate

ggplot(data = graduation, aes(x = Vacant, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +   
 labs(x = "Vacant and Abandoned Residential Properties (%)", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(9)))

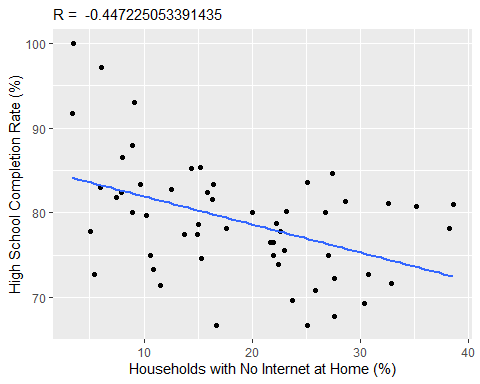
## `geom\_smooth()` using formula = 'y ~ x'



1. Percentage of Households with No Internet at Home vs Completion Rate

ggplot(data = graduation, aes(x = No\_Internet, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +   
 labs(x = "Households with No Internet at Home (%)", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(6)))

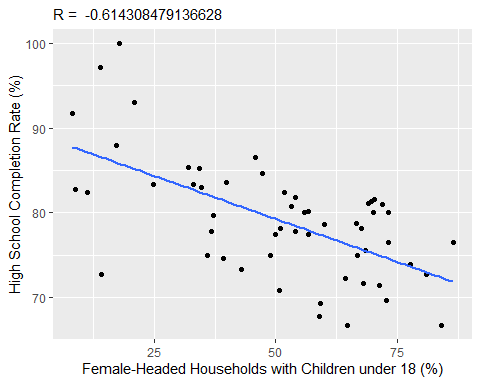
## `geom\_smooth()` using formula = 'y ~ x'



1. Percent of Female-headed Households with Children under 18 vs Completion Rate

ggplot(data = graduation, aes(x = Female\_Led, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +   
 labs(x = "Female-Headed Households with Children under 18 (%)", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(8)))

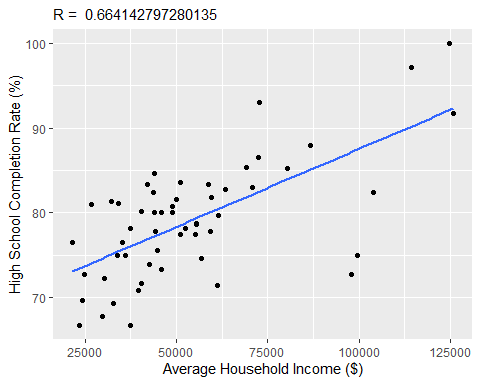
## `geom\_smooth()` using formula = 'y ~ x'



1. household income vs Completion Rate

ggplot(data = graduation, aes(x = Household\_Income, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +   
 labs(x = "Average Household Income ($)", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(11)))

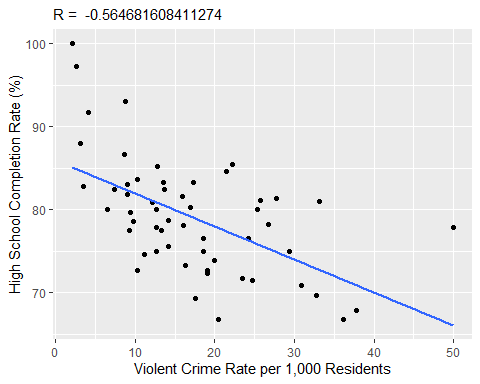
## `geom\_smooth()` using formula = 'y ~ x'



1. Violent Crime vs Completion Rate

ggplot(data = graduation, aes(x = Violent\_Crime, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +   
 labs(x = "Violent Crime Rate per 1,000 Residents", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(10)))

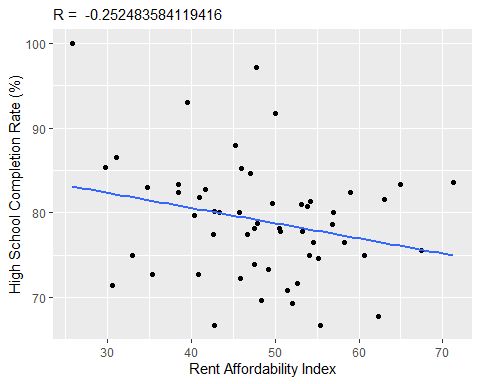
## `geom\_smooth()` using formula = 'y ~ x'



1. Rent affordability vs Completion Rate

ggplot(data = graduation, aes(x = Rent\_Affordability, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +   
 labs(x = "Rent Affordability Index", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(15)))

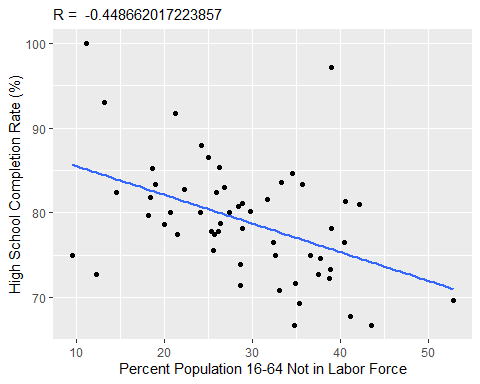
## `geom\_smooth()` using formula = 'y ~ x'



1. Percent not in labor force vs Completion Rate

ggplot(data = graduation, aes(x = Not\_in\_Labor\_Force, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +   
 labs(x = "Percent Population 16-64 Not in Labor Force", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(16)))

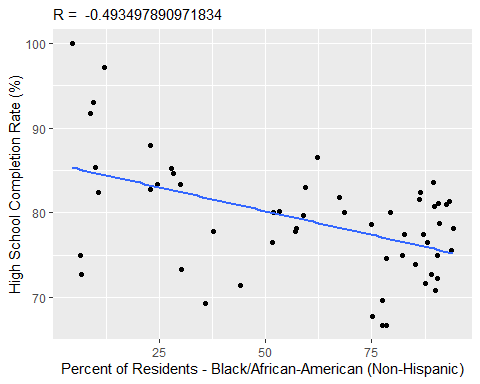
## `geom\_smooth()` using formula = 'y ~ x'



1. Percent Black vs Completion Rate

ggplot(data = graduation, aes(x = Percent\_Black, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +   
 labs(x = "Percent of Residents - Black/African-American (Non-Hispanic)", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(17)))

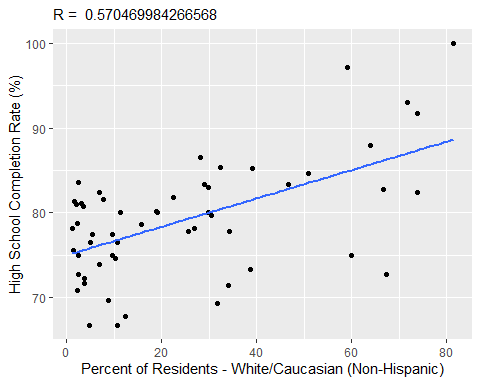
## `geom\_smooth()` using formula = 'y ~ x'



1. Percent White vs Completion Rate

ggplot(data = graduation, aes(x = Percent\_White, y = Completion\_Rate)) +   
 geom\_point() +  
 geom\_smooth(method = lm, se = FALSE) +   
 labs(x = "Percent of Residents - White/Caucasian (Non-Hispanic)", y = "High School Completion Rate (%)", subtitle = paste("R = ", calculate\_R(18)))

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



#5. save plot as png #{r} #ggsave("graduation.png",width = 18,height = 6) #