

How to use GGPlot



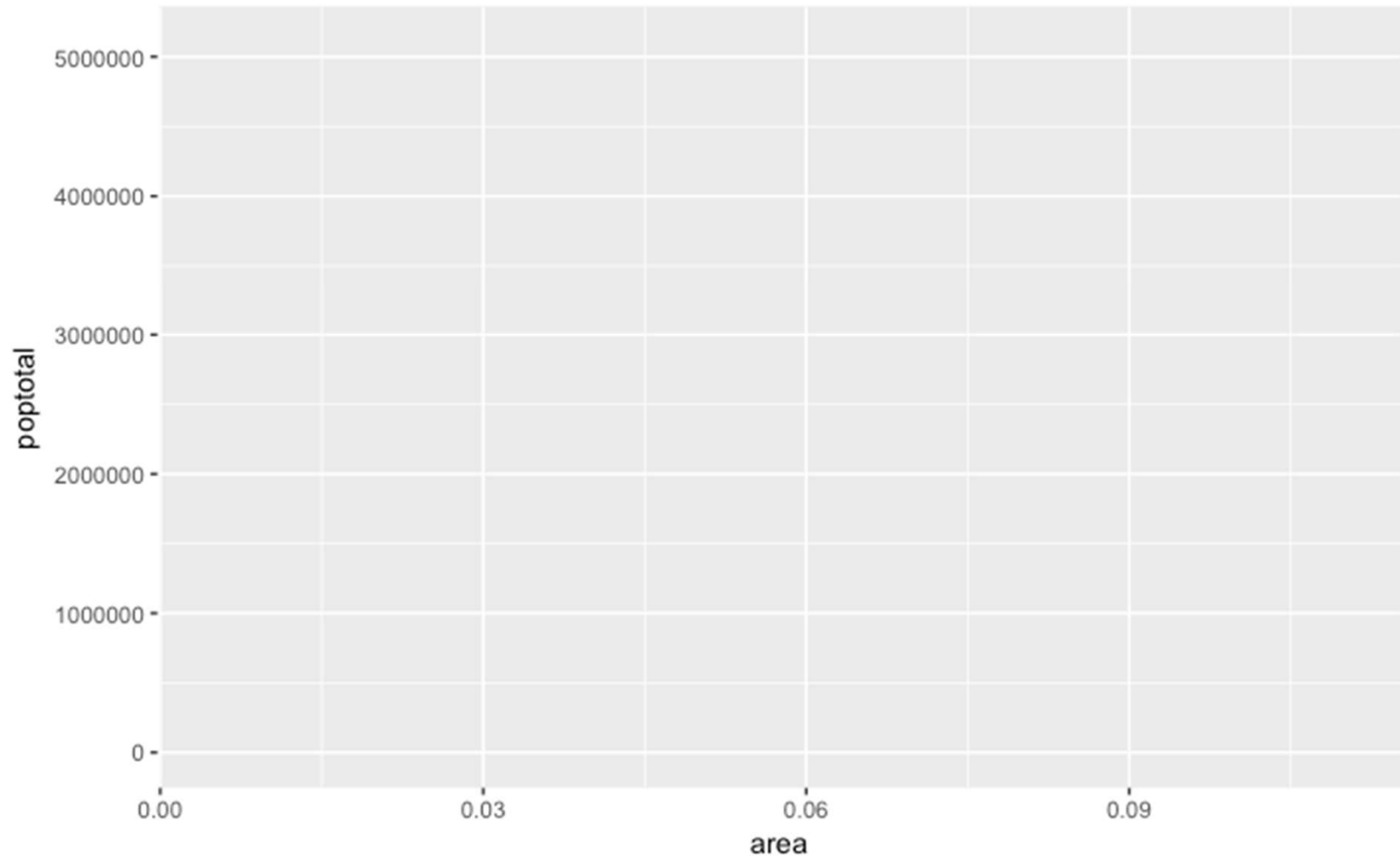
What is ggplot2?

- Graphics framework available in R

The Setup

- `# Setup`
- `options(scipen=999) # turn off scientific notation like 1e+06`
- `library(ggplot2)`
- `data("midwest", package = "ggplot2") # load the data`
- `# midwest <- read.csv("http://goo.gl/G1K41K") # alt source`
- `# Init Ggplot`
- `# area and poptotal are columns in 'midwest'`
- `# aes() function is used to specify the X and Y axes.`
- `ggplot(midwest, aes(x=area, y=poptotal))`

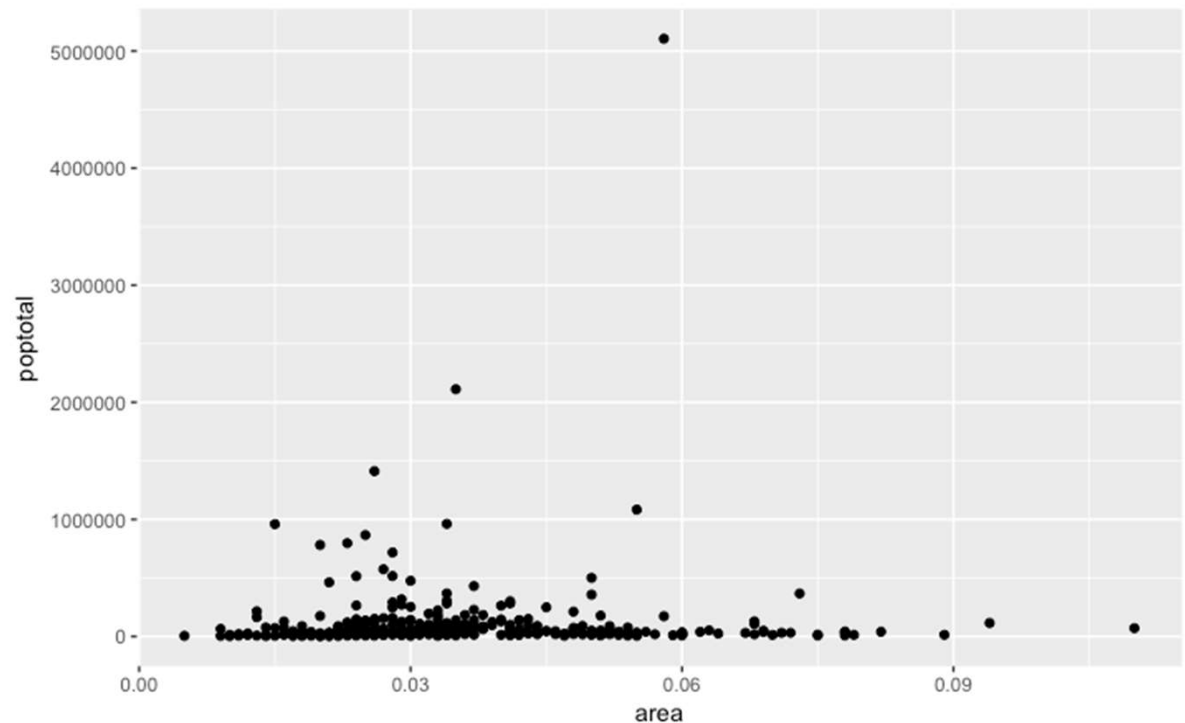
The Setup - Output



How to Make a Simple Scatterplot

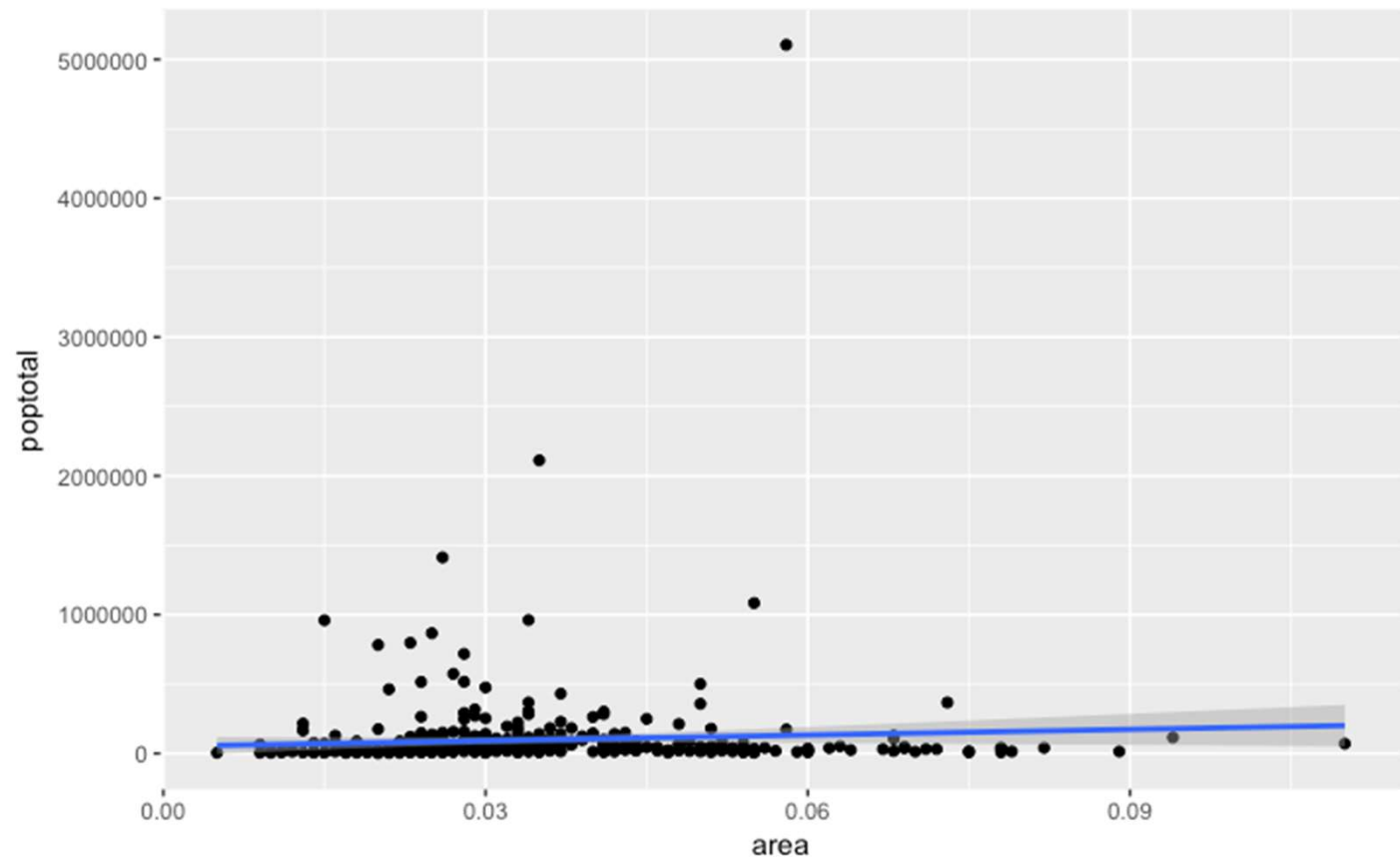
- Let's make a scatterplot on top of the blank ggplot by adding points using a geom layer called `geom_point`.
 - `ggplot(midwest, aes(x=area, y=poptotal)) + geom_point()`
- Like `geom_point()`, there are many such geom layers which we will see.

For now, let's just add a smoothing layer using `geom_smooth`



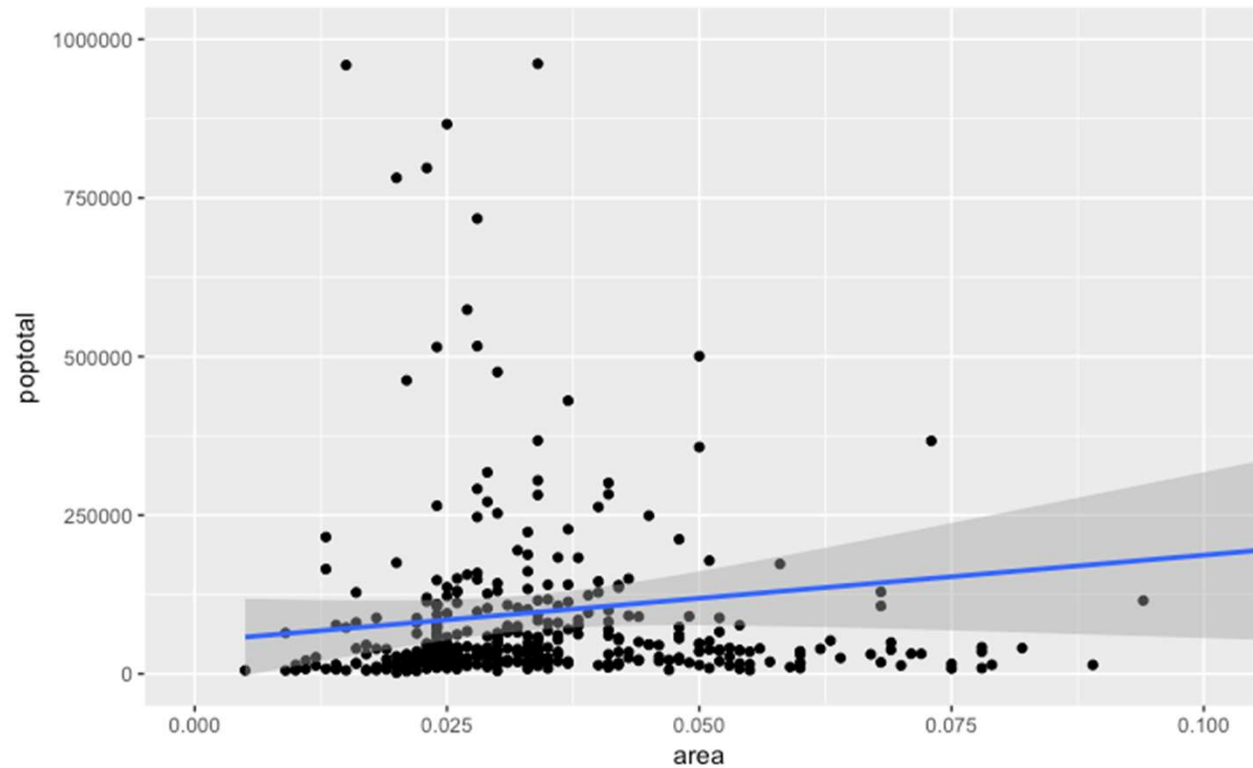
How to Make a Simple Scatterplot

- `g <- ggplot(midwest, aes(x=area, y=poptotal)) + geom_point() + geom_smooth(method="lm") # set se=FALSE to turnoff confidence bands`
- `plot(g)`



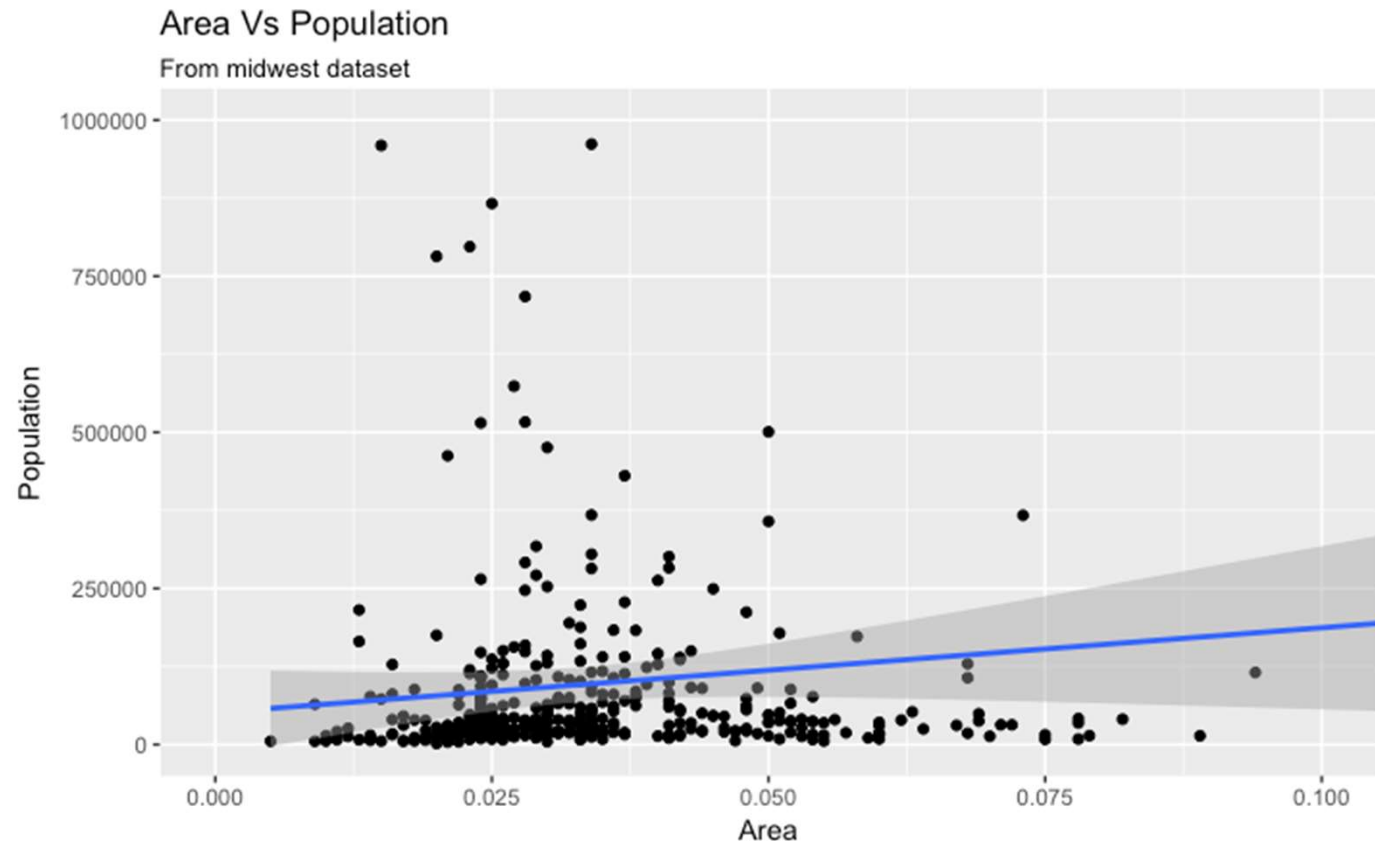
Adjusting the X and Y axis limits

- `g <- ggplot(midwest, aes(x=area, y=poptotal)) + geom_point() + geom_smooth(method="lm")`
- `g1 <- g + coord_cartesian(xlim=c(0,0.1), ylim=c(0, 1000000))` # zooms in
- `plot(g1)`



How to Change the Title and Axis Labels

- `g1 + labs(title="Area Vs Population", subtitle="From midwest dataset", y="Population", x="Area", caption="Midwest Demographics")`



The full function call

- # Full Plot call
- library(ggplot2)
- ggplot(midwest, aes(x=area, y=poptotal)) +
- geom_point() +
- geom_smooth(method="lm") +
- coord_cartesian(xlim=c(0,0.1), ylim=c(0, 1000000)) +
- labs(title="Area Vs Population", subtitle="From midwest dataset", y="Population", x="Area", caption="Midwest Demographics")

How to Change the Color and Size of Points

- `gg <- ggplot(midwest, aes(x=area, y=poptotal)) +`
- `geom_point(aes(col=state), size=3) + # Set color to vary based on state categories.`
- `geom_smooth(method="lm", col="firebrick", size=2) +`
- `coord_cartesian(xlim=c(0, 0.1), ylim=c(0, 1000000)) +`
- `labs(title="Area Vs Population", subtitle="From midwest dataset", y="Population", x="Area", caption="Midwest Demographics")`
- `plot(gg)`

How to Change the Color and Size of Points

