ds1

Matthew Robinson

library(ggplot2)

# Section 3.2.4

# Exercise 1

Run ggplot(data = mpg). What do you see?

ggplot(data = mpg)

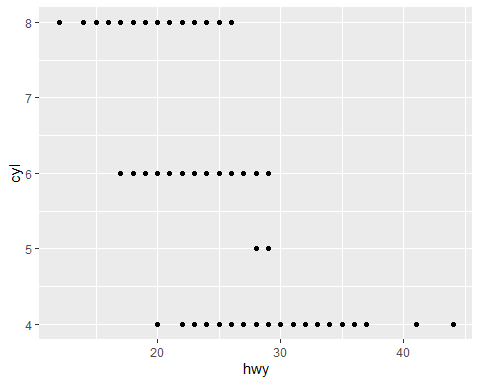


Running this command gives an empty plot because there is no geom function.

# Exercise 4

Make a scatter plot of hwy vs cyl.

ggplot(data = mpg) + geom\_point(mapping = aes(x = hwy, y = cyl))



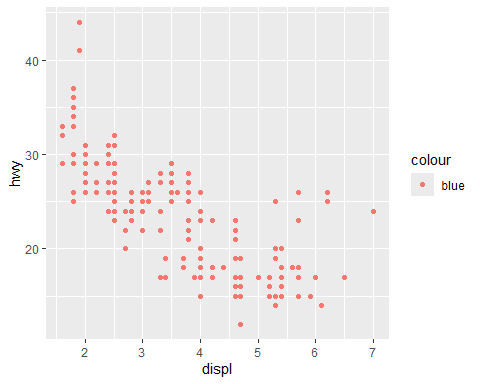
The above plot shows the relationship between hwy and cyl.

# Section 3.3.1

# Exercise 1

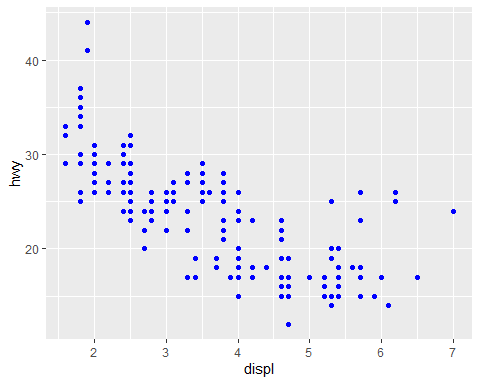
What’s gone wrong with this code? Why are the points not blue?

ggplot(data = mpg) + geom\_point(mapping = aes(x = displ, y = hwy, color = "blue"))



The “color = blue” command should be after the parentheses as shown below.

ggplot(data = mpg) + geom\_point(mapping = aes(x = displ, y = hwy), color = "blue")

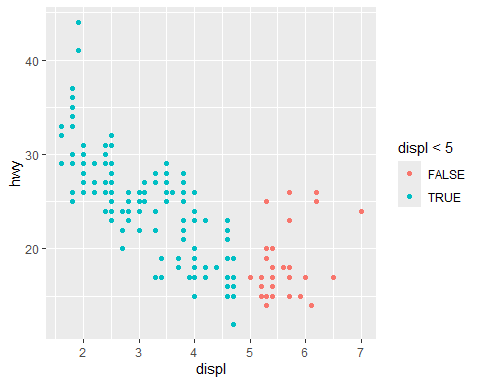


This gives the correct plot with blue points.

# Exercise 6

What happens if you map an aesthetic to something other than a variable name?

ggplot(data = mpg) + geom\_point(mapping = aes(x = displ, y = hwy, color = displ < 5))



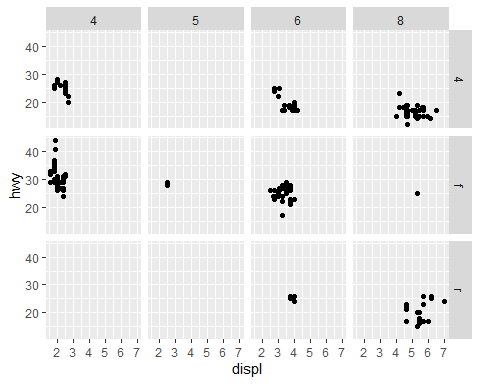
The above example gives different colors to the points depending on if displ is less than 5.

# Section 3.5.1

# Exercise 2

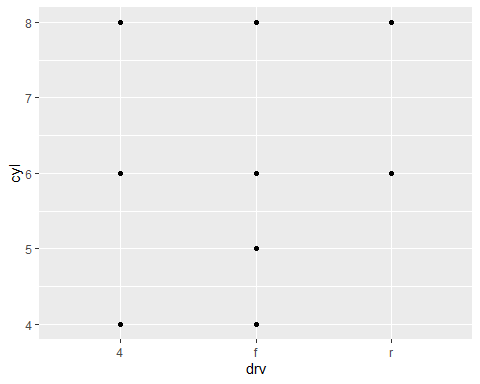
What do the empty cells in plot with facet\_grid(drv ~ cyl) mean?

ggplot(data = mpg) + geom\_point(mapping = aes(x = displ, y = hwy)) + facet\_grid(drv ~ cyl)



The empty cells represent combinations of drv and cyl which do not exist in the data.  
How do they relate to this plot?

ggplot(data = mpg) + geom\_point(mapping = aes(x = drv, y = cyl))

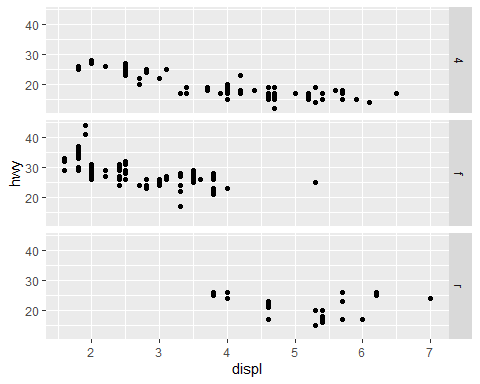


The dots on this plot represent combinations of drv and cyl which do exist in the data.

# Exercise 3

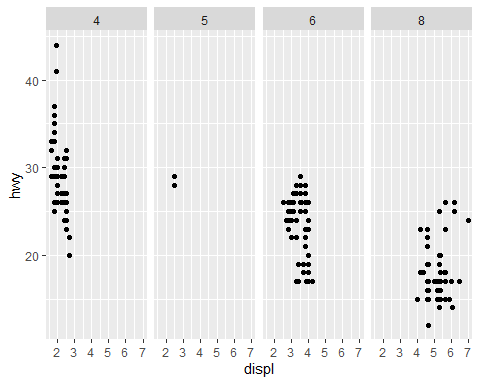
What plots does the following code make? What does . do?

ggplot(data = mpg) + geom\_point(mapping = aes(x = displ, y = hwy)) + facet\_grid(drv ~ .)



The above plot shows the relationship between displ and hwy divided by drv.  
The plot below shows the same relationship divided by cyl instead.

ggplot(data = mpg) + geom\_point(mapping = aes(x = displ, y = hwy)) + facet\_grid(. ~ cyl)



The . makes it to where you only divide the data based on one variable instead of two.

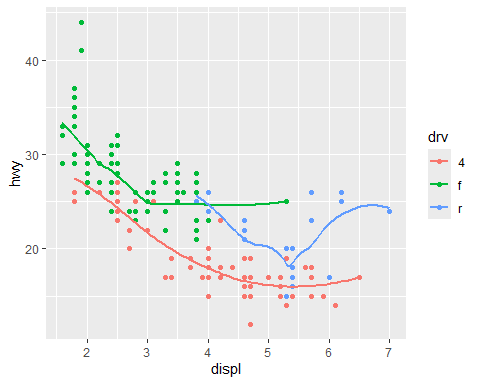
# Section 3.6.1

# Exercise 2

Run this code in your head and predict what the output will look like. Then, run the code in R and check your predictions.

ggplot(data = mpg, mapping = aes(x = displ, y = hwy, color = drv)) + geom\_point() + geom\_smooth(se = FALSE)

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



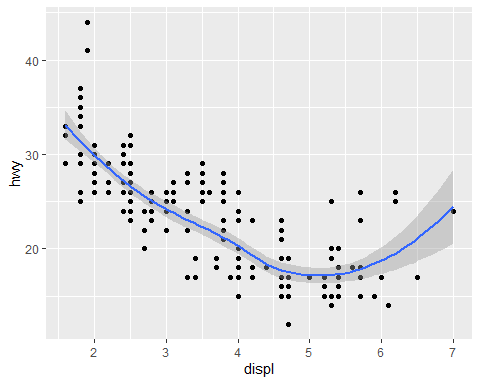
I was expecting the lines to match up with the points, and this prediction was correct.

# Exercise 5

Will these two graphs look different? Why/why not?

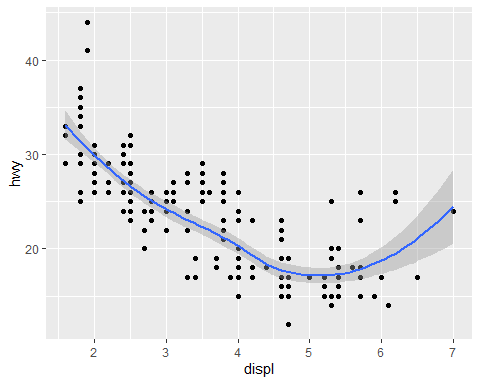
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) + geom\_point() + geom\_smooth()

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



ggplot() + geom\_point(data = mpg, mapping = aes(x = displ, y = hwy)) + geom\_smooth(data = mpg, mapping = aes(x = displ, y = hwy))

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

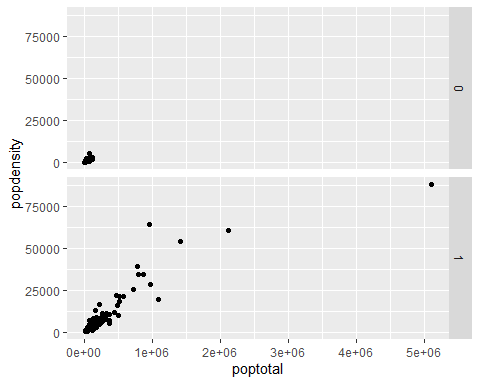


These two graphs do not look different because they are plotting the same thing.

# Midwest Demographics

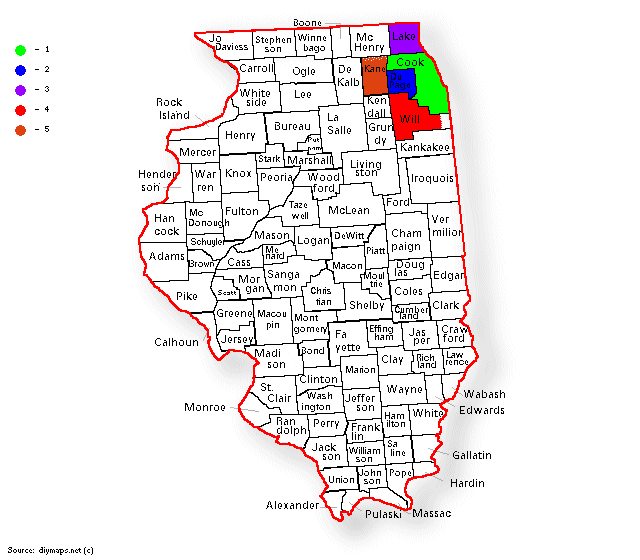
This data set contains data on the demographics of each county in five midwestern states: Illinois, Indiana, Michigan, Ohio, and Wisconsin. The following plot shows the total population and population density of each of these counties divided by state and whether or not they are considered a part of a metropolitan area.

ggplot(data = midwest) + geom\_point(mapping = aes(x = poptotal, y = popdensity)) + facet\_grid(inmetro ~ .)



When looking at this plot, it is clear that the most populated and most densely populated counties are all in metropolitan areas. This makes sense because metropolitan areas are centered around large cities with large populations. If a county is not close enough to one of these cities to be in a metropolitan area, it will probably not have a large population. In fact, the most populous county in each of these states is home to the core city of a metropolitan area. To get a closer look at this, we will look at each of these five states individually. (Please note that the following data is from 2023 and may not match exactly with the previous data.)

# Illinois

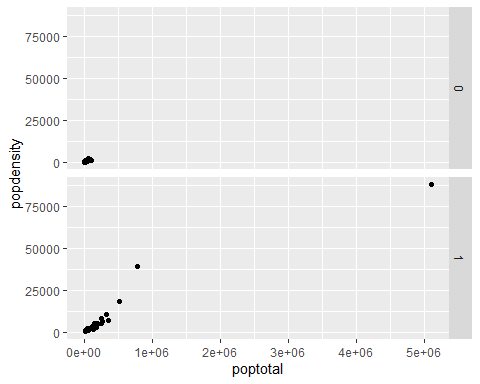


Illinois Counties

| County | County Seat | Metro Area | Population |
| --- | --- | --- | --- |
| Cook | Chicago | Chicago | 5,087,072 |
| DuPage | Wheaton | Chicago | 921,213 |
| Lake | Waukegan | Chicago | 708,760 |
| Will | Joliet | Chicago | 700,728 |
| Kane | Geneva | Chicago | 514,982 |

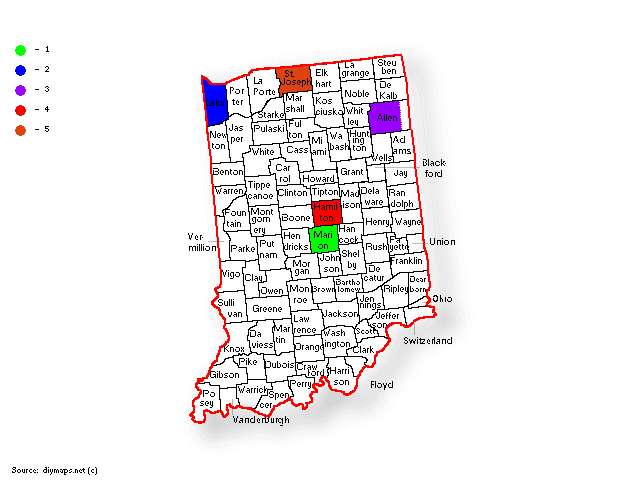
As shown in the above table, the top five most populous counties in Illinois are all part of the Chicago Metropolitan Area. This is the only one of these five states to have all of the top five largest counties all be in the same metropolitan area. This makes sense as Chicago is the third largest city in the country and the largest city in the midwest.

ggplot(data = dplyr::filter(midwest, state == "IL")) + geom\_point(mapping = aes(x = poptotal, y = popdensity)) + facet\_grid(inmetro ~ .)



In the above plot, the point to the far right represents Cook County. DuPage and Lake counties are also separated from the cluster. Will and Kane counties can be seen at the edge of the cluster.

# Indiana

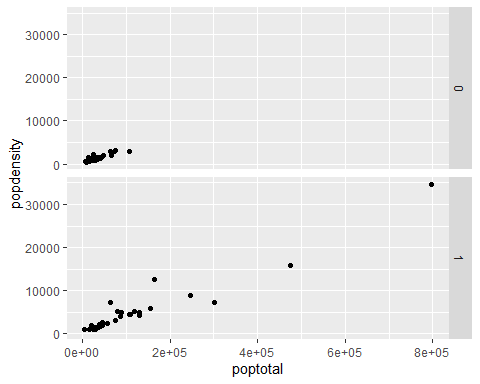


Indiana Counties

| County | County Seat | Metro Area | Population |
| --- | --- | --- | --- |
| Marion | Indianapolis | Indianapolis | 968,460 |
| Lake | Crown Point | Chicago | 500,598 |
| Allen | Fort Wayne | Fort Wayne | 394,545 |
| Hamilton | Noblesville | Indianapolis | 371,645 |
| St. Joseph | South Bend | South Bend | 272,848 |

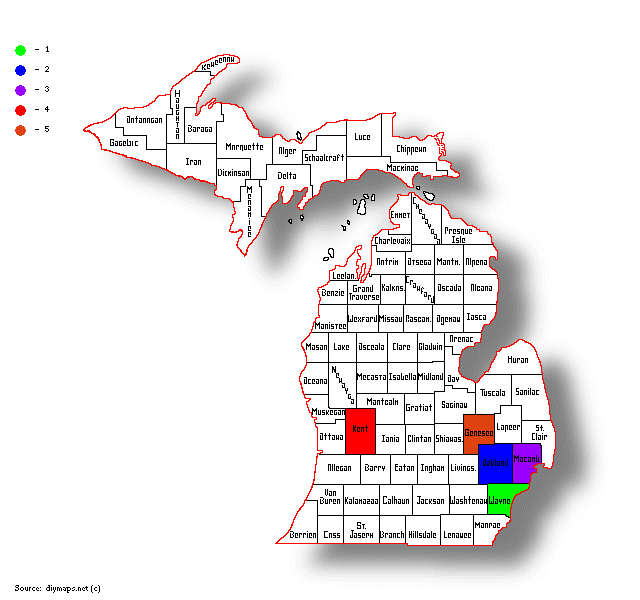
As shown in the above table, only two of the top five most populous counties in Indiana share a metropolitan area. Marion and Hamilton counties are both in the Indianapolis Metropolitan Area which makes sense as Indianapolis is the capital and largest city in Indiana. Interestingly, Lake County is the only county in the 25 counties listed to be a part of a metropolitan area anchored in a different state. It makes sense that it would be Chicago for the reasons listed earlier along with its close proximity to Indiana.

ggplot(data = dplyr::filter(midwest, state == "IN")) + geom\_point(mapping = aes(x = poptotal, y = popdensity)) + facet\_grid(inmetro ~ .)



In the above plot, the point to the far right represents Marion County. Lake County is also well removed from the cluster. Allen, Hamilton, and St. Joseph counties are closer to the cluster but are still separated from it.

# Michigan

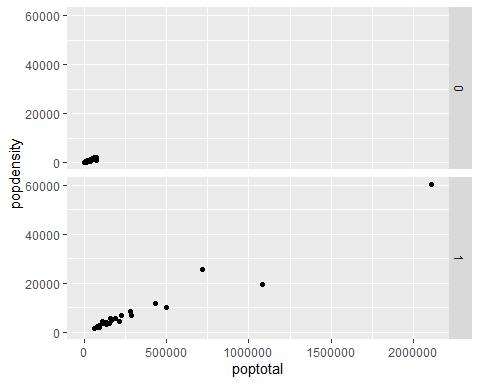


Michigan Counties

| County | County Seat | Metro Area | Population |
| --- | --- | --- | --- |
| Wayne | Detroit | Detroit | 1,751,169 |
| Oakland | Pontiac | Detroit | 1,270,426 |
| Macomb | Mt. Clemens | Detroit | 875,101 |
| Kent | Grand Rapids | Grand Rapids | 661,354 |
| Genesee | Flint | Flint | 401,522 |

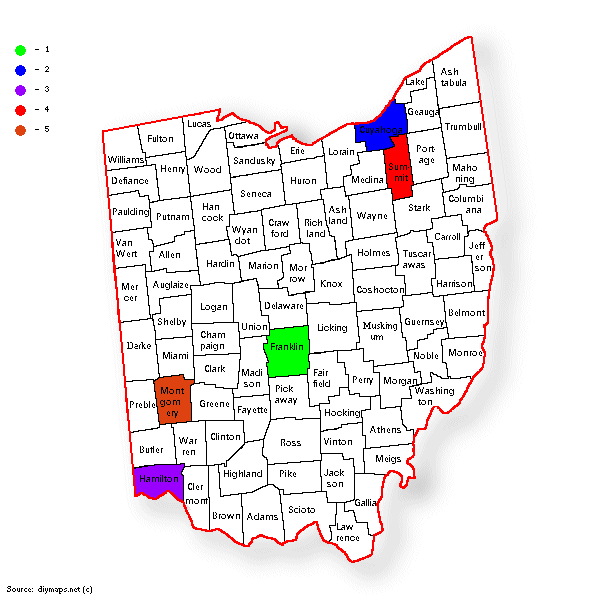
As shown in the above table, the top three most populous counties in Michigan are all part of the Detroit Metropolitan Area. This makes sense as it is the largest city in the state. Since Flint Metropolitan Area is a part of the Detroit Combined Statistical Area, Kent County is the only county on this list outside of Greater Detroit.

ggplot(data = dplyr::filter(midwest, state == "MI")) + geom\_point(mapping = aes(x = poptotal, y = popdensity)) + facet\_grid(inmetro ~ .)



In the plot above, the point to the far right represents Wayne county. Oakland and Macomb counties are also well removed from the cluster. Kent and Genesee counties are closer to the cluster but are still separated from it.

# Ohio

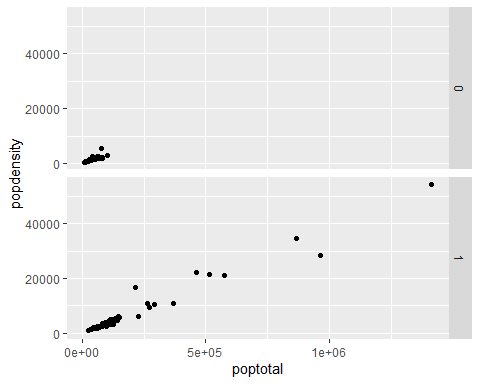


Ohio Counties

| County | County Seat | Metro Area | Population |
| --- | --- | --- | --- |
| Franklin | Columbus | Columbus | 1,326,063 |
| Cuyahoga | Cleveland | Cleveland | 1,233,088 |
| Hamilton | Cincinnati | Cincinnati | 827,058 |
| Summit | Akron | Akron | 535,733 |
| Montgomery | Dayton | Dayton | 533,796 |

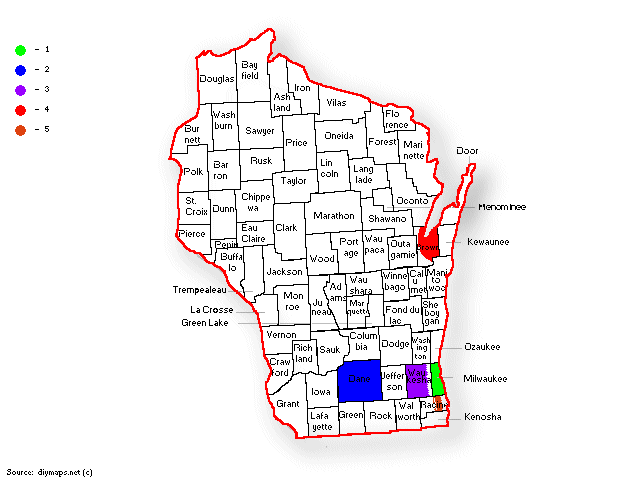
As shown in the above table, the top five most populous counties in Ohio are all part of separate metropolitan areas. This is the only one of these five states to not have any of its top five most populous counties be in the same metropolitan area. This makes sense as Ohio has several major cities in different parts of the state. However, the Akron Metropolitan Area is a part of the Cleveland Combined Statistical Area.

ggplot(data = dplyr::filter(midwest, state == "OH")) + geom\_point(mapping = aes(x = poptotal, y = popdensity)) + facet\_grid(inmetro ~ .)



In the plot above, the point to the far right represents Franklin County. Cuyahoga and Hamilton counties are also well removed from the cluster. Summit and Montgomery counties are in a group with Lucas County (home of Toledo, not listed above) closer to the cluster.

# Wisconsin

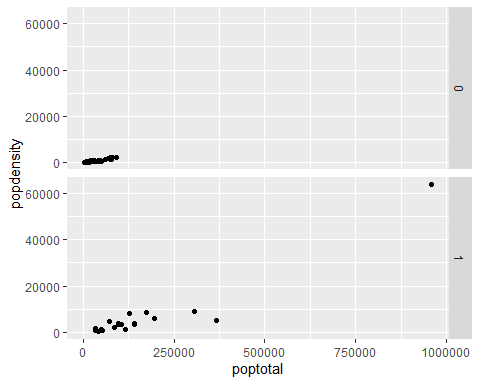


Wisconsin Counties

| County | County Seat | Metro Area | Population |
| --- | --- | --- | --- |
| Milwaukee | Milwaukee | Milwaukee | 916,205 |
| Dane | Madison | Madison | 575,347 |
| Waukesha | Waukesha | Milwaukee | 412,591 |
| Brown | Green Bay | Green Bay | 271,417 |
| Racine | Racine | Racine | 196,613 |

As shown in the above table, only two of the top five most populous counties in Wisconsin share a metropolitan area. Milwaukee and Waukesha counties are both in the Milwaukee Metropolitan Area. This makes sense as Milwaukee is the largest city in the state. Furthermore, the Racine Metropolitan Area is a part of the Milwaukee Combined Statistical Area.

ggplot(data = dplyr::filter(midwest, state == "WI")) + geom\_point(mapping = aes(x = poptotal, y = popdensity)) + facet\_grid(inmetro ~ .)



In the plot above, the point to the far right represents Milwaukee County. Dane and Waukesha counties are also separated from the cluster. Brown and Racine counties can be seen on the edge of the cluster.

# Conclusion

Based on this data, it is clear that the most populous and most densely populated counties are in metropolitan areas. While some states are dominated by one metropolitan area such as Illinois with Chicago, others have more spread out populations such as Ohio. Furthermore, the most populous county in each of these states is home to the core city of a metropolitan area. While it may seem like common sense that population will be greater in metropolitan areas, it is interesting to see how the data confirms that.