**Name:**

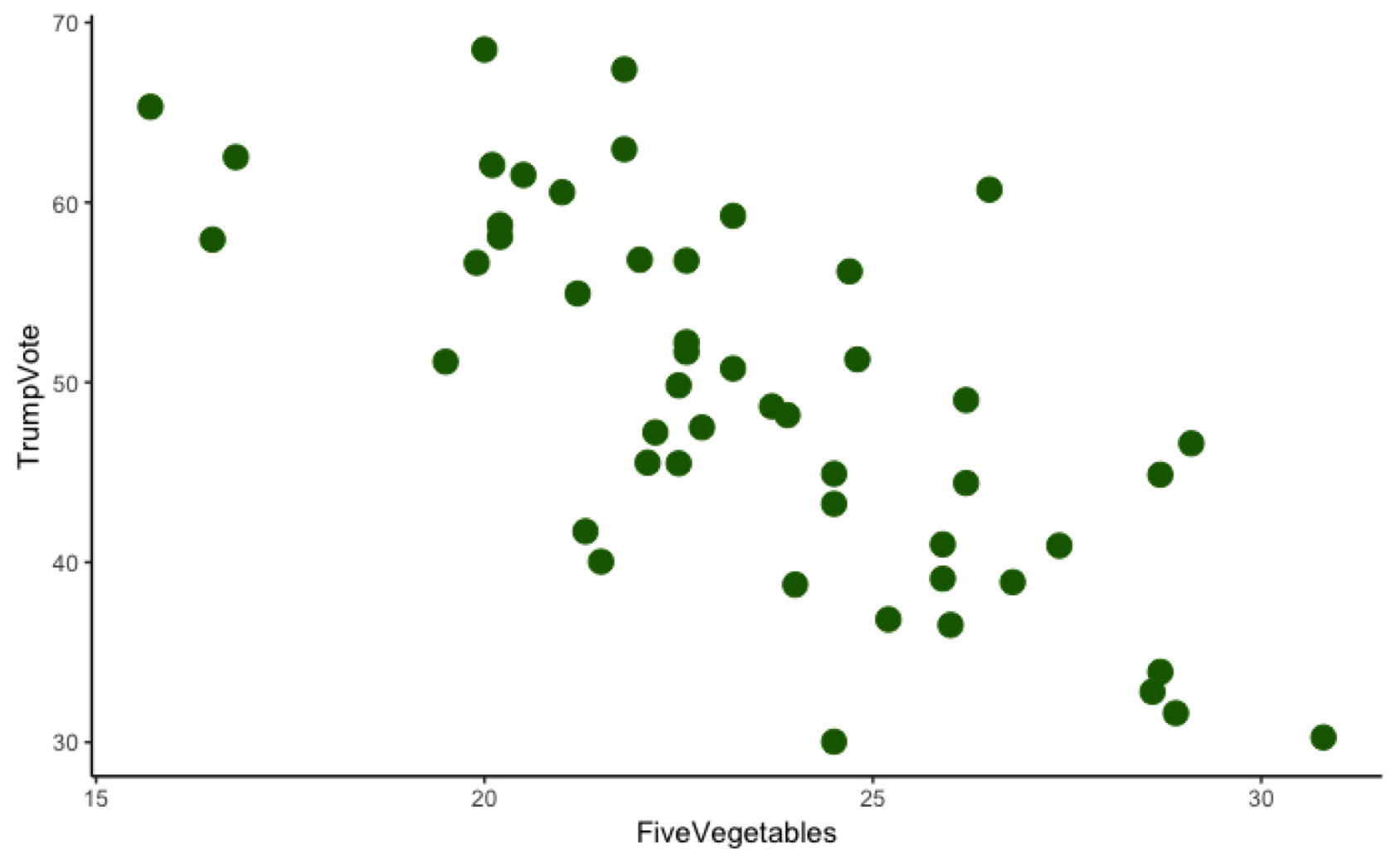
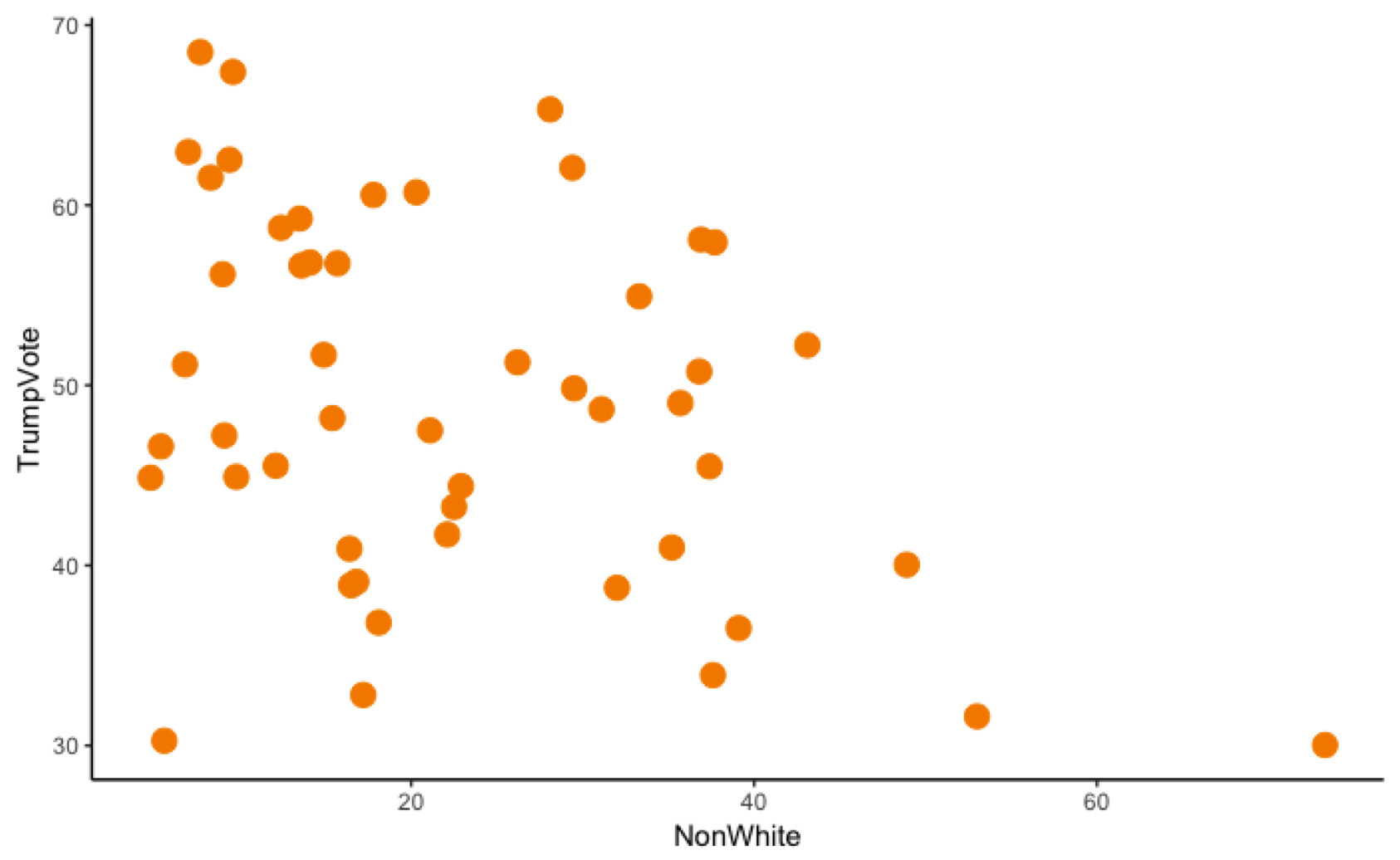
**Classwork 7**

**US States: Trump, Vegetables, and NonWhite People**

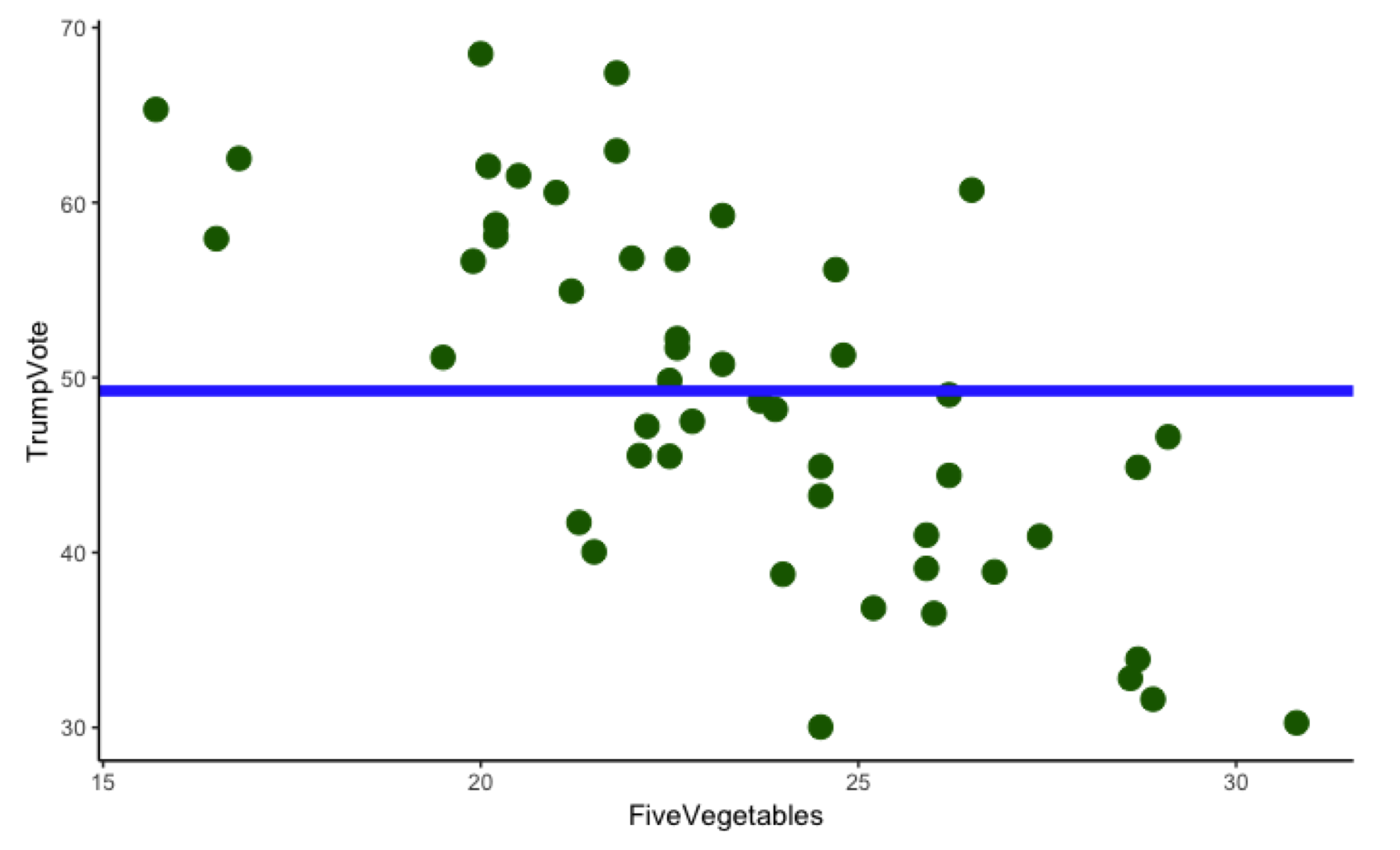
1. Let’s put in **TrumpVote** into the **USStates** data frame. What will this bit of code do? Can I put the numbers in any order I want?

USStates$TrumpVote <- c(62.08, 51.28, 48.67, 60.57, 31.62, 43.25, 40.93, 41.72, 49.02, 50.77, 30.03, 59.26, 38.76, 56.82, 51.15, 56.65, 62.52, 58.09, 44.87, 33.91, 32.81, 47.5, 44.92, 57.94, 56.77, 56.17, 58.75, 45.5, 46.61, 41, 40.04, 36.52, 49.83, 62.96, 51.69, 65.32, 39.09, 48.18, 38.9, 54.94, 61.53, 60.72, 52.23, 45.54, 30.27, 44.41, 36.83, 68.5, 47.22, 67.4)

1. Let’s take a look at the distribution of **TrumpVote** in a point plot. What do you notice? (Think shape, center, spread.)
2. Which do you think explains more the variation in **TrumpVote**: **FiveVegetables** or **NonWhite**? Why? Write both ideas as a word equation.
3. How would you create a visualization to explore these two word equations? Write the R code here.
4. As you eyeball the visualizations, which variable seems to explain *more* variation: **FiveVegetables** or **NonWhite**? What aspect of the visualizations are you looking at to make that judgment?
5. If we knew that a state had a large proportion of healthy eaters, how would we adjust our prediction of **TrumpVote**?
6. If we knew that a state had a large proportion of nonwhite residents, how would we adjust our prediction of **TrumpVote**?
7. If we didn’t know anything about a state, what should we predict their **TrumpVote** to be?
8. How would we add the mean and median into the scatterplots that we made already? (Write the additional R code.) Sketch the lines here.

1. Why is the mean and median represented as lines? Why not dots?
2. Why is the mean and median the same on both graphs? Is that just a coincidence?
3. If the median really is “the middle,” how many dots would be above and below this line?
4. If the mean really is “the middle,” what does it balance? Draw the stuff that the mean balances in the plot below (just draw a few so we get the idea). What are these called?



1. Can you prove the median really is the “middle” (in its own special way) by using R?
2. Can you prove the mean really is the “middle” (in its own special way) by using R?
3. There is an election coming up in 2020. We could use a very simple model, just the mean, to predict how a random state might vote in the next year. What are the pros and cons of using this simple model?