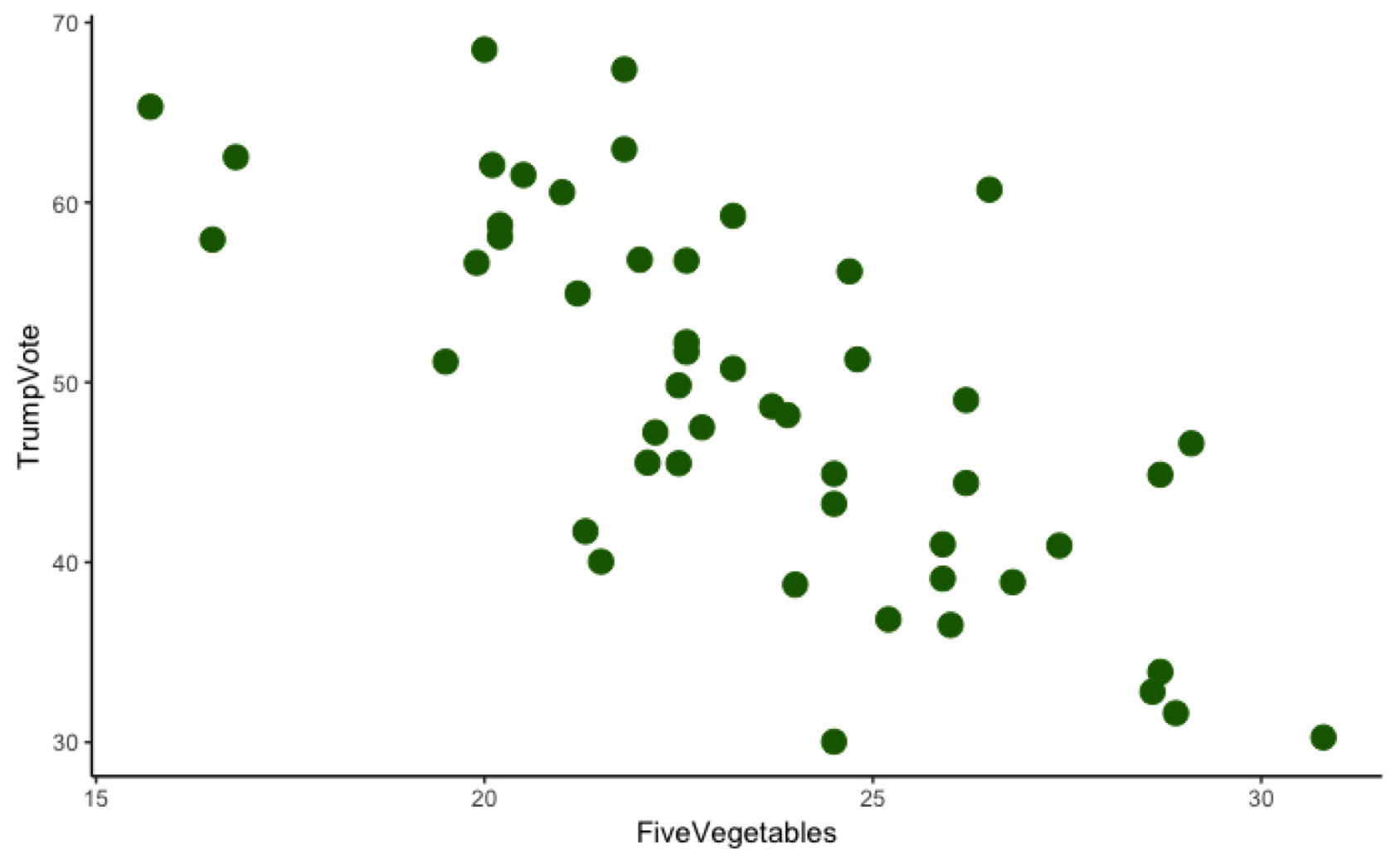
**Name:**

**Classwork 8**

**US States: Statistical Error in TrumpVote**

1. If we used the mean as an empty model, what would it predict is the **TrumpVote** for each state? What R function would we use to get those values? What R code would you use to save those predictions in the **USStates** data frame?
2. If we put plotted the empty model’s prediction on our scatterplot, where would those predictions go? How would we do that in R?



1. Why do they all go there? (Go back and re-consider Question 10 in Classwork 7.)
2. If we used the mean as a model, it is clearly not a great model. Only a handful of states had **TrumpVote** values that were close to the mean. How would we figure out how much error there was for each state? What R function would you use? How would you write the R code to save the residuals for each state into the data frame **USStates**?
3. Go back to the drawing in question 2. How would you draw the residuals?
4. How much total error is here? If we summed up the residuals, what would that total be? Does that mean that this empty model is perfect?
5. How would we write the empty model in General Linear Model (GLM) notation? Let’s also modify our word equation for randomness to align better with this notation.
6. What use is the empty model?
7. If we shuffled **FiveVegetables** in this data frame, would we estimate a different empty model? Would the empty model change?
8. Why is the empty model (GLM notation: ) a stand-in for DGP of *randomness*?