Name: Audrey Pangallo

UmichID: 33021701

What dataset are you working with: Bob Ross, Obviously.

List 3 questions that you can ask with your dataset.

Q1: How often do happy little trees appear in Bob Ross’s seasons?

Q2: Does Bob Ross appear to have a preference for majestic mountains, or tranquil swaying palm trees?

Q3: Is he more of a sun or moon guy?

List the associated null hypothesis for each question:

Q1: Happy little trees make equal appearances in all Bob Ross seasons

Q2: Bob Ross does not favor majestic mountains or tranquil palm trees  
Q3: Neither does he favor the sun or the moon.

What statistical test(s) will you use to answer each of the questions:

Q1-3: Given the way they’ve entered this data, it all appears to be categorical (either it’s present or it isn’t). So I think I’d be using a chi-squared test for all of them.

Make a visual plot showing the relationship that you will analyze statistically (e.g. boxplot for t-test or ANOVA; scatterplot for regression; table for chi-square).

Q1: Table for happy trees and season

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

0 1 0 1 1 4 2 3 2 4 2 1 3 0 2 2 3 4 2

1 12 13 12 12 9 11 10 11 9 11 12 10 13 11 11 10 9 11

19 20 21 22 23 24 25 26 27 28 29 30 31

0 1 2 3 2 4 1 4 2 1 2 3 3 1

1 12 11 10 11 9 12 9 11 12 11 10 10 12

Q2: 0 1

0 295 9

1 99 0

Do your data meet the assumptions required for the statistical test you want to run? Please state the assumptions you examined and whether or not your data meet those assumptions:

Q1: the assumptions for a chi squared test are: the sample is randomly selected, that my observations are independent of one another, and that my contingency table is filled out. The contingency table is filled out, see above.

Run the statistical test! Put your results here:

Q1: Pearson's Chi-squared test

data: bob\_ross$trees and bob\_ross$season

X-squared = 23.301, df = 30, p-value = 0.8027

Q2: Pearson's Chi-squared test with Yates' continuity

correction

data: bob\_ross$mountains and bob\_ross$palm\_trees

X-squared = 1.7953, df = 1, p-value = 0.1803

Interpret your results!

Q1: It doesn’t look like Bob Ross favors trees during any of his seasons. That makes sense. That guy loves trees.

Q2: again, there appears to be no significance.

I’ve also gotten errors for both of these and I’m not sure why… I have a feeling it has something to do with the way I’m asking the question but right now my flu ridden brain is having trouble figuring out what the problem is.

Ok what I did was look up what fivethirtyeight did with this data, and it appears they used the 0,1 data (not present, present) to calculate percentages and probabilities.

<https://fivethirtyeight.com/features/a-statistical-analysis-of-the-work-of-bob-ross/>

and then they produced cluster graphs. So in r, I would have to get those probabilities from the raw data somehow, and then use, maybe a z test? I don’t think chi-squared is what I should be using after reading this.