**PART I**

**Distribution of Chosen Variable**



Figure 1. Histogram indicating frequency within breast cancer death rates.



Figure 2. A scatter plot illustrating individual rates for all 335 points. Five breaks were identified and circled.

For this assignment, I decided to examine breast cancer rates (only reported for females). The rates, provided within the data, is the death rate due to breast cancer per 100,000 people in a population. A visual inspection was done in Excel and the output charts are seen in Figures 1 and 2. Majority of the rates are between 18.2 to 29.8%, the most falling between 24 to 27%. The distribution in Figure 1 was calculated automatically, resulting in 14 separate classes. In Figure 2, all 335 points were ordered in descending order. The number of breaks were identified based on intuition, separating the data based on distinct decreases in rates.

**Using CLASSIT to Analyze Data Distribution**

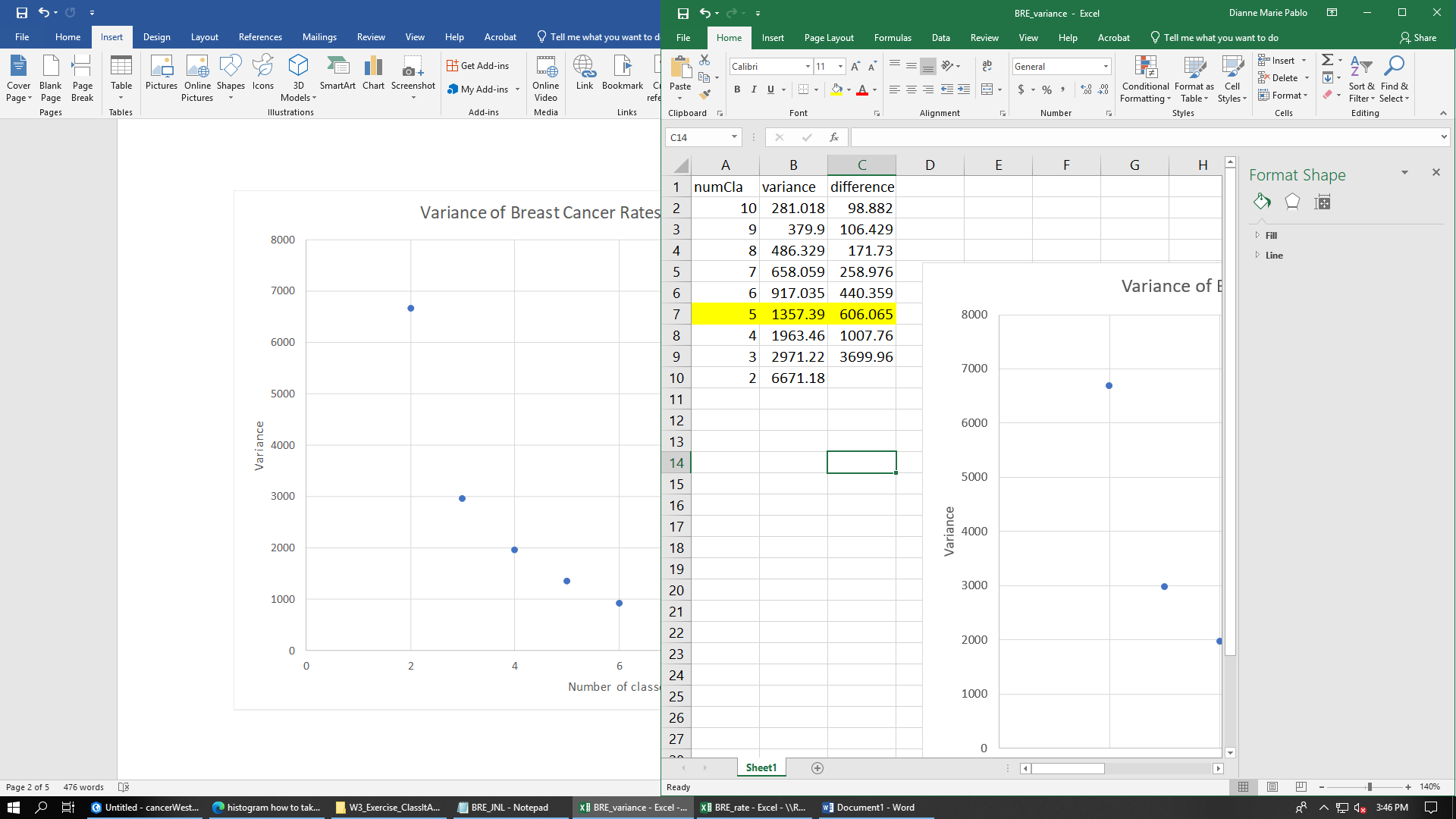
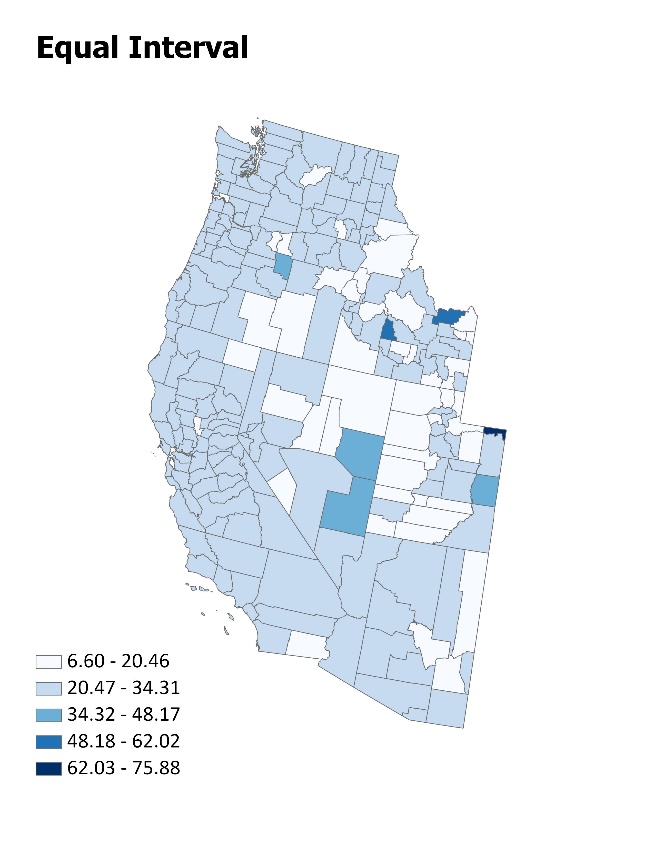
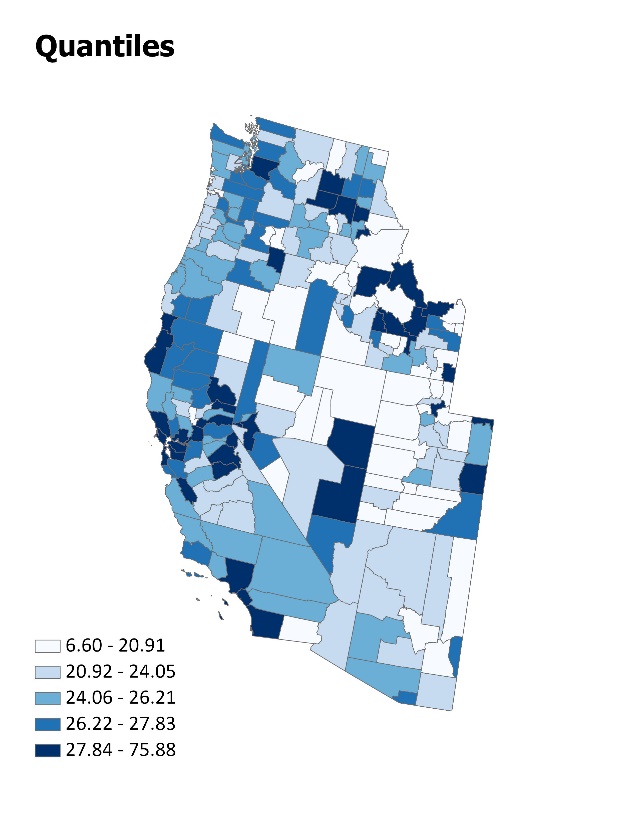
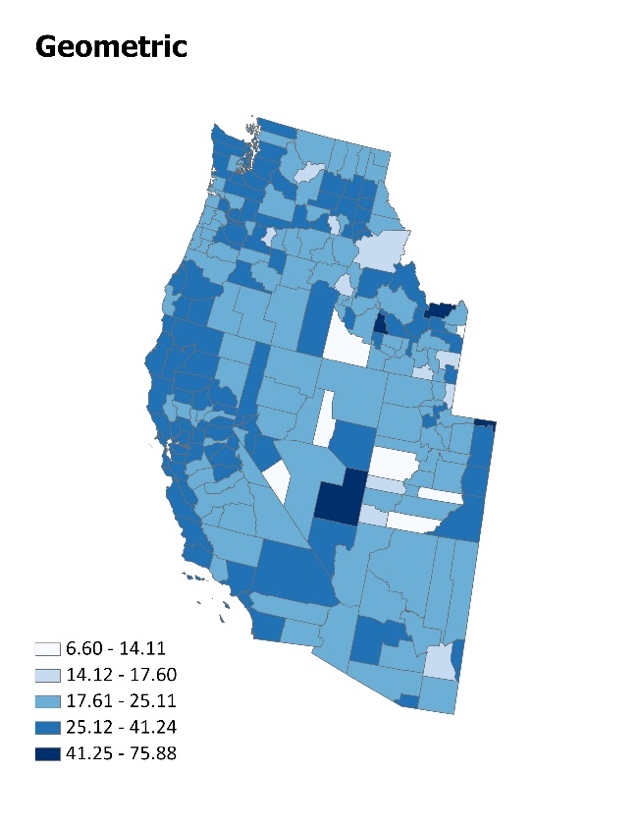


Figure 3. Results from CLASSIT. An elbow is identified at 5 classes (circled, on the left). Variance data is also shown (right) from 2-10 classes, and the difference from the previous number (e.g., 98.882 is the difference between 10 classes and 9 classes, and so on).

Based on the outputs from CLASSIT, the optimal number of classes for this dataset is 5. As seen on the right of Figure 3, the elbow was identified at this point. The variance data also portray that after five classes, the differences in variance between classes decrease insignificantly. The full output dataset from CLASSIT is attached at the end of this report.

**PART 2**

**Choropleth Mapping**

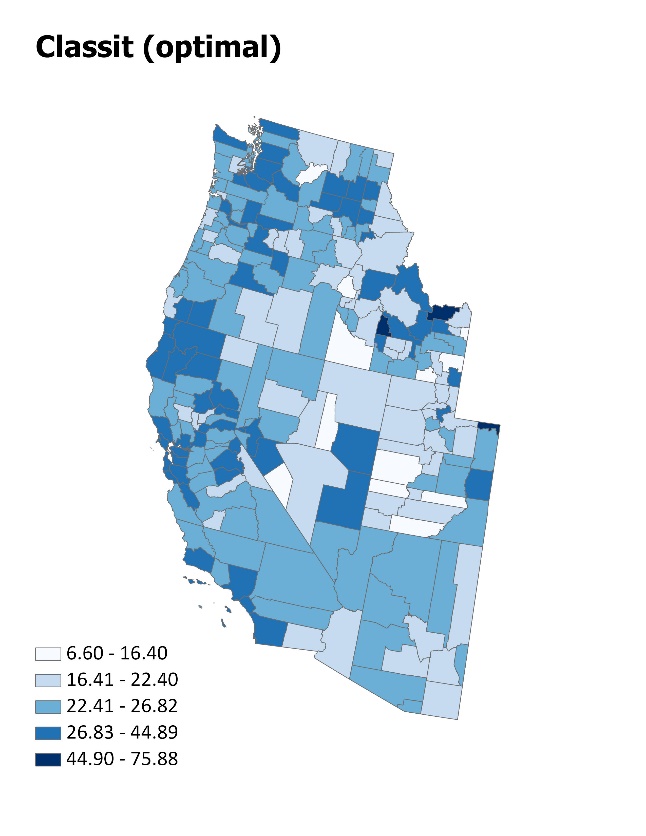


Figure 4. Breast cancer death rates among females in western US counties. Four different classification schemes were used to illustrate class distribution.

**Map Analysis**

Out of all four schemes, the CLASSIT classification is the most visually and statistically realistic. This is because the class breaks were identified using variance data. The equal interval places too much emphasis on the low rates. There are only a few counties in the mid-range class and only one with the highest rate, which is not accurate based on the distribution of the data. With quantile classification, breast cancer death rates between 20.92 and 27.83% are thinly classified. Thus, the last classification in the legend – those above 27.84% - are grouped together. That major difference in rate, in addition to the poorly classified mid-range rates, place counties in the higher end of the distribution, which is also not accurate. Contrastingly, the geometric classification does a poor job in creating breaks between 17.61 to 41.24%. Here, majority of the counties are uniformly classed in this range, which would explain the homogeneity in most locations. Before mapping, it is very important to consider the data and statistically test them in order to determine class breaks. This way, the map is not skewed towards one end of the dataset and readers are not making inaccurate inferences.

**CLASSIT Output**

OPTIMAL CLASSES AS MEASURED BY VARIANCE

A 10 CLASS MAP WITH VARIANCE OF 281.018

Class nobs largest smallest mean variance

10 1 75.880 75.880 75.880 0.000

9 2 57.830 57.220 57.525 0.186

8 1 44.890 44.890 44.890 0.000

7 17 36.430 31.990 32.992 33.649

6 44 30.750 27.800 28.826 27.504

5 85 27.760 25.470 26.744 35.904

4 85 25.320 22.650 24.078 46.347

3 61 22.400 19.680 21.171 35.395

2 31 19.540 14.970 18.071 41.523

1 8 14.110 6.600 11.023 60.455

A 9 CLASS MAP WITH VARIANCE OF 379.900

Class nobs largest smallest mean variance

9 1 75.880 75.880 75.880 0.000

8 2 57.830 57.220 57.525 0.186

7 1 44.890 44.890 44.890 0.000

6 19 36.430 30.600 32.748 43.269

5 98 29.980 26.390 27.834 86.812

4 95 26.280 23.460 24.862 68.964

3 72 23.310 20.190 21.802 58.114

2 39 20.060 14.970 18.439 62.132

1 8 14.110 6.600 11.023 60.455

A 8 CLASS MAP WITH VARIANCE OF 486.329

Class nobs largest smallest mean variance

8 1 75.880 75.880 75.880 0.000

7 3 57.830 44.890 53.313 106.614

6 19 36.430 30.600 32.748 43.269

5 98 29.980 26.390 27.834 86.812

4 95 26.280 23.460 24.862 68.964

3 72 23.310 20.190 21.802 58.114

2 39 20.060 14.970 18.439 62.132

1 8 14.110 6.600 11.023 60.455

A 7 CLASS MAP WITH VARIANCE OF 658.059

Class nobs largest smallest mean variance

7 1 75.880 75.880 75.880 0.000

6 3 57.830 44.890 53.313 106.614

5 23 36.430 29.980 32.267 68.583

4 120 29.770 25.650 27.364 130.713

3 101 25.590 21.930 23.883 93.838

2 78 21.810 16.210 19.843 183.895

1 9 14.970 6.600 11.461 74.307

A 6 CLASS MAP WITH VARIANCE OF 917.035

Class nobs largest smallest mean variance

6 3 75.880 57.220 63.643 224.791

5 20 44.890 30.600 33.355 183.322

4 122 29.980 25.720 27.478 150.661

3 103 25.660 21.930 23.917 100.011

2 78 21.810 16.210 19.843 183.895

1 9 14.970 6.600 11.461 74.307

A 5 CLASS MAP WITH VARIANCE OF 1357.394

Class nobs largest smallest mean variance

5 3 75.880 57.220 63.643 224.791

4 20 44.890 30.600 33.355 183.322

3 149 29.980 24.820 27.054 273.760

2 128 24.670 19.450 22.385 293.949

1 35 19.300 6.600 16.299 381.472

A 4 CLASS MAP WITH VARIANCE OF 1963.459

Class nobs largest smallest mean variance

4 4 75.880 44.890 58.955 488.554

3 143 36.430 25.650 28.153 663.073

2 141 25.590 20.190 23.090 323.963

1 47 20.060 6.600 17.177 487.741

A 3 CLASS MAP WITH VARIANCE OF 2971.217

Class nobs largest smallest mean variance

3 4 75.880 44.890 58.955 488.554

2 210 36.430 23.550 26.976 1304.147

1 121 23.480 6.600 20.033 1178.303

A 2 CLASS MAP WITH VARIANCE OF 6671.176

Class nobs largest smallest mean variance

2 4 75.880 44.890 58.955 488.554

1 331 36.430 6.600 24.438 6182.370

A 1 CLASS MAP WITH VARIANCE OF 11380.063

Class nobs largest smallest mean variance

1 335 75.880 6.600 24.850 11379.776