PAPER #901

United States of Obesity

Mary Richardson, richamar@gvsu.edu
Grand Valley State University

Diann Reischman, reischmd@gvsu.edu
Grand Valley State University

Paul Stephenson, stephenp@gvsu.edu
Grand Valley State University

Key Words: Introductory statistics; Thematic map; Regression; Correlation

Abstract

Maps are the preferred method used to display statistical data coming from a geographic frame. Maps are aesthetically appealing and are helpful in identifying geographic patterns in a data set. In this paper we will present an example of a choropleth map and we will discuss how we use the map to illustrate regression and correlation in an introductory statistics course.

1. Some Background on Thematic Maps

A thematic map is a map of a theme or topic. There are numerous varieties of thematic maps. Most of the maps utilized by media outlets such as a map that shows the status of sales taxes by state or a map that shows the world population density are thematic maps. A thematic map displays the spatial pattern of a theme or series of attributes. Thematic maps emphasize spatial variation of one or a small number of geographic distributions. These distributions may be physical phenomena such as climate or human characteristics such as population density and health issues. Location, of course, is important to provide a reference base of where selected phenomena are occurring.

A choropleth map is a special case of a thematic map. Choropleth maps are particularly suited for charting phenomena that are evenly distributed within each set area. Raw data, for example,

a population distribution, should not be mapped with a choropleth map. However; if a derived value can be obtained from raw data (such as population densities), then the choropleth map can apply. Choropleth maps are one of the simplest methods of representing data that have been collected for areal units. They are commonly used to map census demographic data based on townships or census divisions. To produce a choropleth map, the observations are grouped into a set of classes based on their data values and then each class is shaded with an appropriate color.

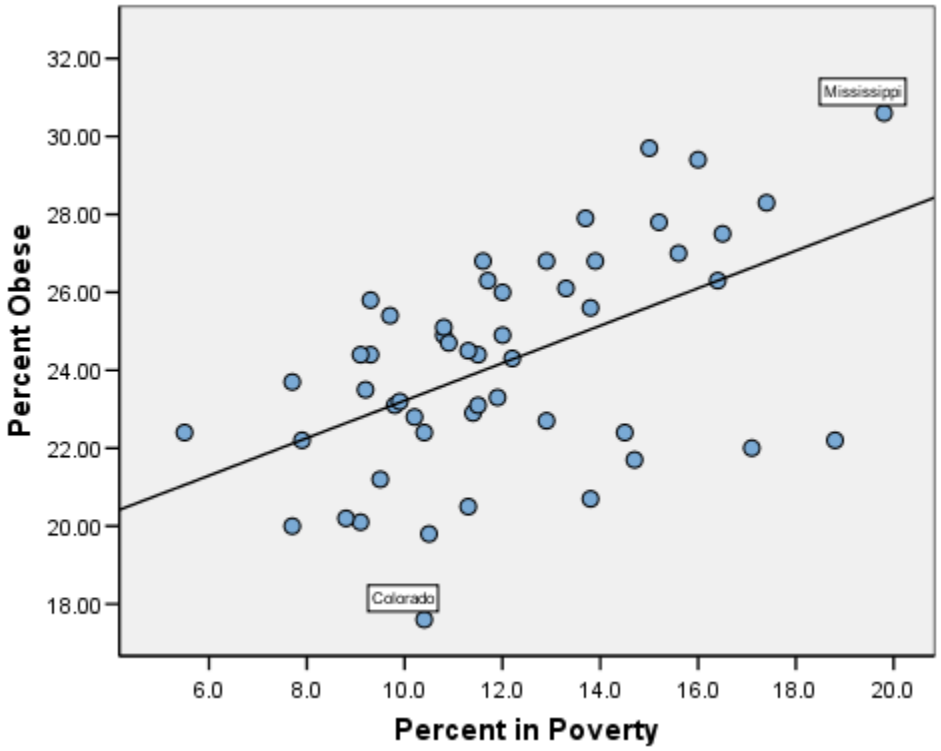
2. The Activity

The activity "United States of Obesity" has students explore descriptive calculations for the percentages of obese adults in the United States (based upon 3-year averages from 2004 to 2006). Students refer to the choropleth "United States of Obesity" map (on the Activity Worksheet in the Appendix) and calculate measures of center and spread for the obesity percentages. The mean percent obese is 24.22%. The standard deviation is 2.80%. The minimum is 17.6%, Quartile 1 is 22.4%, the median is 24.4%, Quartile 3 is 26.3%, and the maximum is 30.6%. In order to check for outliers, the interquartile range (IQR) is calculated as: Q3 - Q1 = 26.3% - 22.4% = 3.9%. So, a STEP, or 1.5(IQR), is 1.5(3.9%) = 5.85%. The fences are obtained by taking a 'step' below Quartile 1 and a 'step' above Quartile 3. Thus, the lower fence is: 22.4% - 5.85% = 16.55%. And the upper fence is: 26.3% + 5.85% = 32.15%. There are no obesity percentages outside of the fences, so there are no outliers.

A histogram is provided to display the distribution of the obesity percentages. Students use the numerical calculations, along with the histogram to give a complete description of the distribution of the obesity percentages. Additionally, students comment on the use of the histogram vs. the map to display the distribution. The advantage of using a histogram to display the distribution is that we can easily see that the obesity percentages range from about 18% to about 31%. The distribution appears to be roughly mound-shaped, with the highest frequency of percentages between about 23% and 25%. There are no outliers. The advantage of using a map to display the distribution is that we can see differences in the obesity percentages over different regions of the United States. We see that the highest obesity percentages appear to be in the extreme South. We also see that the West tends to have lower obesity percentages than the East.

We provide students with a data table that displays State Poverty Rates (3-Year Averages, 2004-2006, in percentages). The Activity Worksheet contains the data table. We use the data table, along with the map to have students investigate whether there is a linear relationship between a state's poverty percentage and obesity percentage. Students construct a scatterplot using a state's poverty percentage as the explanatory variable and a state's obesity percentage as the response variable. Figure 1 shows the scatterplot.

Figure 1. Scatterplot of state poverty percentage vs. state obesity percentage.



Students interpret the scatterplot. As the percent in poverty increases, the percent obese also tends to increase. On the map it was noted that Colorado was the state with the lowest percent obese, while Mississippi had the highest percent. These states have been marked on the scatterplot.

Students determine the estimated regression equation for predicting a state's obesity percentage from the state's poverty percentage and plot the regression equation on the scatterplot. The equation is $\hat{y} = 18.413 + .481x$. Students interpret the slope. On the average, we estimate that for every increase of 1 percentage point in a state's poverty rate, the state's obesity rate increases by .48 percentage points. The numerical value of the correlation coefficient between a state's obesity percentage and a state's poverty rate is given and an interpretation is provided. The correlation coefficient is $r = .522$, indicating that there is a moderately strong linear association between these two variables.

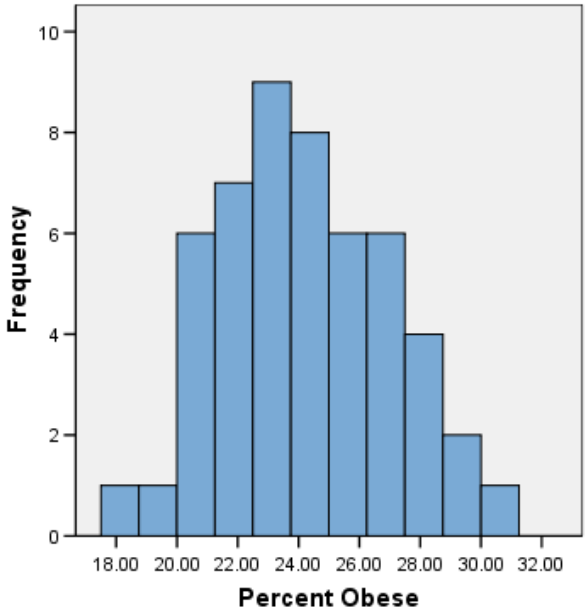
3. Conclusions

Maps are frequently used to display data distributions that have a geographic context. Maps are commonly seen in newspapers, magazines, on web pages, and on TV newscasts. In fact, maps are now as commonly and sometimes more commonly seen as data displays typically discussed in an introductory statistics course such as pie charts and bar graphs. We propose the use and interpretation of statistical maps in an introductory course. And, in this paper, we provided one example of how a statistical map can be used in the classroom.

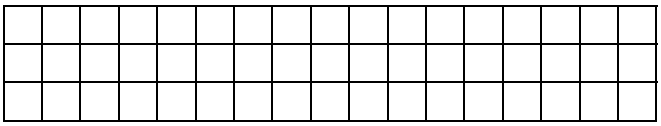
Appendix Activity Worksheet

1. Refer to the "United States of Obesity" map. Calculate the following for the obesity percentages: mean, standard deviation, and five-number summary. In addition, determine if any of the state obesity percentages are outliers.

2. Below is a histogram that displays the distribution of the obesity percentages. Use the numerical calculations above, along with the histogram to give a complete description of the distribution of the obesity percentages. Additionally, comment on the use of the histogram vs. the map to display the distribution of the obesity percentages.



3. Refer to the data table that displays state poverty rates. Construct a scatterplot with each state's obesity percentage on the vertical axis and each state's corresponding poverty rate on the horizontal axis.

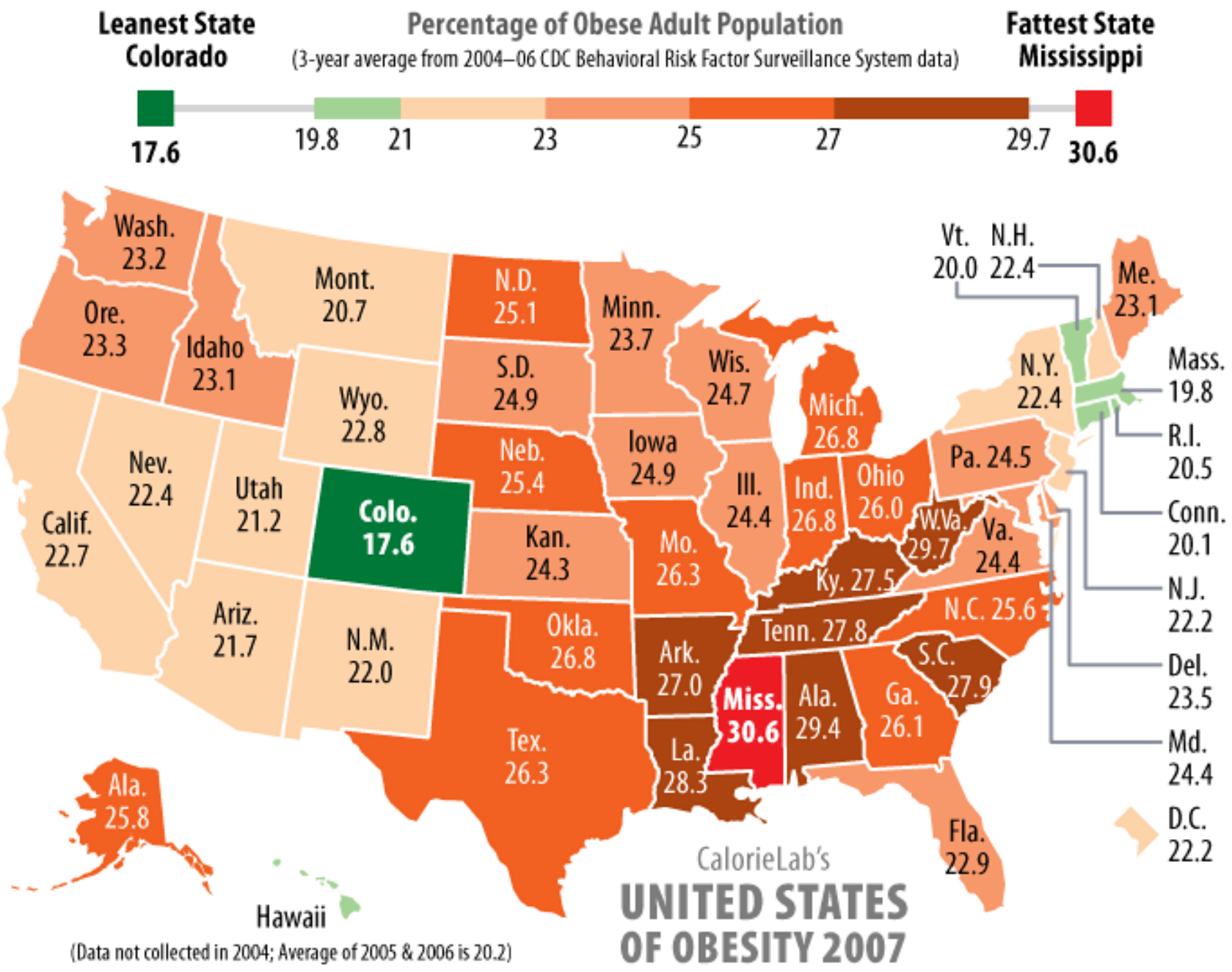


4. Interpret the scatterplot.

5. Determine the estimated regression equation for predicting a state's obesity percentage from the state's poverty rate. Plot the regression equation on the scatterplot.

6. Interpret the slope of the estimated regression equation in the context of this problem.

7. Give the numerical value of and interpret the correlation coefficient between a state's obesity percentage and a state's poverty rate.


Leanest State
Colorado
Percentage of Obese Adult Population
(3-year average from 2004–06 CDC Behavioral Risk Factor Surveillance System data)
Fattest State
Mississippi
17.6
19.8
21
23
25
27
29.7
30.6
Wash. 23.2
Ore. 23.3
Idaho 23.1
Mont. 20.7
Wyo. 22.8
Nev. 22.4
Calif. 22.7
Utah 21.2
Colo. 17.6
Ariz. 21.7
N.M. 22.0
N.D. 25.1
S.D. 24.9
Neb. 25.4
Kan. 24.3
Okla. 26.8
Tex. 26.3
Minn. 23.7
Iowa 24.9
Mo. 26.3
Ark. 27.0
La. 28.3
Wis. 24.7
Ill. 24.4
Mich. 26.8
Ind. 26.8
Ohio 26.0
Ky. 27.5
Tenn. 27.8
Miss. 30.6
Ala. 29.4
Vt. 20.0
N.H. 22.4
Me. 23.1
N.Y. 22.4
Pa. 24.5
W.Va. 29.7
Va. 24.4
N.C. 25.6
S.C. 27.9
Ga. 26.1
Fla. 22.9
Mass. 19.8
R.I. 20.5
Conn. 20.1
N.J. 22.2
Del. 23.5
Md. 24.4
D.C. 22.2
Ala. 25.8
Hawaii
(Data not collected in 2004; Average of 2005 & 2006 is 20.2)
CalorieLab's
UNITED STATES
OF OBESITY 2007

State Poverty Rates: 3-Year Averages, 2004-2006 (in percentages)

State	Rate	State	Rate
Alabama	16.0	North Carolina	13.8
Alaska	9.3	North Dakota	10.8
Arizona	14.7	Ohio	12.0
Arkansas	15.6	Oklahoma	13.9
California	12.9	Oregon	11.9
Colorado	10.4	Pennsylvania	11.3
Connecticut	9.1	Rhode Island	11.3
Delaware	9.2	South Carolina	13.7
D. C.	18.8	South Dakota	12.0
Florida	11.4	Tennessee	15.2
Georgia	13.3	Texas	16.4
Hawaii	8.8	Utah	9.5
Idaho	9.8	Vermont	7.7
Illinois	11.5	Virginia	9.1
Indiana	11.6	Washington	9.9
Iowa	10.8	West Virginia	15.0
Kansas	12.2	Wisconsin	10.9
Kentucky	16.5	Wyoming	10.2
Louisiana	17.4		
Maine	11.5		
Maryland	9.3		
Massachusetts	10.5		
Michigan	12.9		
Minnesota	7.7		
Mississippi	19.8		
Missouri	11.7		
Montana	13.8		
Nebraska	9.7		
Nevada	10.4		
N. Hampshire	5.5		
New Jersey	7.9		
New Mexico	17.1		
New York	14.5		

References

Background on thematic maps adapted from the following sources:
http://chnm.gmu.edu/worldhistorysources/unpacking/mapsmain.html
http://en.wikipedia.org/wiki/Thematic_map_
http://www.fes.uwaterloo.ca/crs/geog165/maps.htm
http://en.wikipedia.org/wiki/Geographic_information_system
http://en.wikipedia.org/wiki/Map_making
http://www.phil.uni-passau.de/histhw/tutcarto/english/index-frames-en.html
http://www.ncjrs.gov/html/nij/mapping/

Data Source: http://www.irp.wisc.edu/faqs/faq3/table2.htm

United States of Obesity Map Source: http://calorielab.com/news/2007/08/06/fattest-states-2007/