Appendices

Bibliography
Methodological and Technical Notes
A. BIBLIOGRAPHY


Centers for Disease Control and Prevention, Quitting Smoking. (2017, February 1). Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/cessation/cessation/quitting/

Centers for Disease Control and Prevention, Reproductive and Birth Outcomes. (2016, August 31). Retrieved from https://ephtracking.cdc.gov/showRbLBWGrowthRetardationEnv.action


Johns Hopkins University. (2014, August 6). *Stroke risk, related death rates have declined over past 20 years, study shows*. Retrieved from Johns Hopkins University, The Hub: https://hub.jhu.edu/2014/08/06/stroke-rate-decline-study/


Rubin, P. R. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55. Retrieved from https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/biomet/70/1/10.1093/biomet/70.1.41/2/70-1-41.pdf?Expires=1488584930&Signature=LoZP4uTuVqlcq3GL5p9gWzzPdcx04rglNIsOm-w-HTDofm0JqS7wO1Anq08LM4c97FmZ1BkiOoBiW985mWIadxTdlBuVHZi6OQZE57eMJq3d5


Substance Abuse and Mental Health Services Administration. (2013). Retrieved from Report to Congress on the Nation's Substance Abuse and Mental Health Workforce Issues:


http://ww5.komen.org/BreastCancer/DisparitiesInBreastCancerScreening.html


https://www.healthlandscape.org/Brigh_Spots_Appropriat/Diabetes_Preventive_Care.pdf


B. METHODOLOGICAL AND TECHNICAL NOTES

Definition of a County

The data in this report come from more than 15 secondary data sources. Different institutions organize their data into a particular universe of spatial units based on their analytical needs. The county-level data files provided by these sources—or sub-county data aggregated from source files—were organized to meet the Appalachian Regional Commission’s standard representation of the United States as consisting of 3,113 counties, which adheres to the Bureau of Economic Analysis’s county-level delineation of the nation. However, many of the data sources disaggregate the country into 3,143 “county units” (with some slight variation around this number). The data from many of these secondary data sources thus have to be converted into the 3,113-county universe, and this is done by combining several county-level units. The most frequent example of this takes place in Virginia, where independent cities are combined with surrounding counties in order to meet the ARC/BEA organizational structure. If source data provided numerators and denominators, these values were used to compute figures such as rates and percentages for each indicator. When only computed figures were provided, a weighting variable from another source (such as the 2014 American Community Survey (ACS) population figures) was used to create a weighted average of values. In rare cases, data from source counties were distributed to more than one of the 3,113 counties. This was most often necessary for five Alaska boroughs/county-equivalents, which were recently reallocated from three county-equivalents. In these cases, values from the three county-equivalents were directly assigned to the new areas based on predominately geographic overlap.

Mapping Procedures

The measures included in this report are examined at the county level, with national and regional maps for each indicator displaying its variation across counties. For ease of interpretation, the maps for each variable were designed to have darker colors represent values indicating poor health status, or values for social or behavioral factors that contribute to poor health status.

The value groupings used to determine the color categories in both the national and regional maps are determined by the national quintiles of each respective indicator: five groupings with equal numbers of constituent counties in each category (with some variation due to rounding, clustering, and suppressed data values). The quintiles are determined by using the breaks at the 20th, 40th, 60th, and 80th percentiles of the national distribution. Since both the national and regional maps use these national groupings, the national maps will have the same number of counties in each category. The maps for the Appalachian Region, however, will not necessarily have an equal distribution among categories and colors, as the quintiles are based on the national data. For example, if the Appalachian Region as a whole performs worse than the national average for any particular measure, it is likely that more than 20 percent of the Region’s counties will be shaded in the darkest color on the map.

---

8 The number of counties may vary by one between groups. For example, there are 3,113 counties analyzed, which does not divide equally into groups of five, meaning for indicators with complete data, three groups will have 623 and two will have 622.
Age Adjusting Mortality Rates

Age adjusting mortality rates allows for comparisons among counties with different age distributions. Counties with greater numbers of elderly residents can generally expect higher mortality rates than counties with less elderly residents. Thus, a county with higher unadjusted (crude) mortality rates, which suggest poor health, may actually be relatively healthy but simply have a larger number of older residents (and thus a higher overall baseline risk of death). Using data from the Compressed Mortality Files from CDC, we compared the distributions of county populations by age cohort to the standard population distribution for the country as a whole. Using these population distributions by age cohort as a base, the mortality rates in this report are age adjusted and standardized based on the Year 2000 Standard Million Population (see Table 71). This provides the reader with the ability to accurately compare mortality across counties with different age distributions. Only infant mortality is not age adjusted in this report, because the age distribution of a population is not relevant to the measure. The YPLL indicator is obtained directly from County Health Rankings & Roadmaps, which age-adjusted YPLL prior to publishing.

Table 71: Year 2000 Standard Million Population for the United States

<table>
<thead>
<tr>
<th>Age</th>
<th>2000 Standard Population Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 1 year</td>
<td>13,818</td>
</tr>
<tr>
<td>1-4 years</td>
<td>55,317</td>
</tr>
<tr>
<td>5-9 years</td>
<td>72,533</td>
</tr>
<tr>
<td>10-14 years</td>
<td>73,032</td>
</tr>
<tr>
<td>15-19 years</td>
<td>72,169</td>
</tr>
<tr>
<td>20-24 years</td>
<td>66,478</td>
</tr>
<tr>
<td>25-34 years</td>
<td>135,573</td>
</tr>
<tr>
<td>35-44 years</td>
<td>162,613</td>
</tr>
<tr>
<td>45-54 years</td>
<td>134,834</td>
</tr>
<tr>
<td>55-64 years</td>
<td>87,247</td>
</tr>
<tr>
<td>65-74 years</td>
<td>66,037</td>
</tr>
<tr>
<td>75-84 years</td>
<td>44,841</td>
</tr>
<tr>
<td>85 years and over</td>
<td>15,508</td>
</tr>
<tr>
<td>All Years</td>
<td>1,000,000</td>
</tr>
</tbody>
</table>

Source: https://wonder.cdc.gov/wonder/help/cmf.html

The formulas to convert crude rates to weighted rates by age cohort and total population age-adjusted are:

\[
(Deaths / pop) \times 100,000 = CRUDE RATE \\
CRUDE RATE \times \left(\frac{\text{standard pop in each age cohort}}{1,000,000}\right) = WEIGHTED RATE \\
\text{Sum (WEIGHTED RATES) all cohorts} = \text{AGE-ADJUSTED RATE for total population}
\]

In this example, the crude data are reported in increments of 100,000 residents.
For example, consider the crude and age-adjusted mortality rates in two states: Utah and Maine, the former of which has a relatively younger population. Table 72 displays the crude and age adjusted rates using two sources: CDC Wonder, an interactive web tool that allows users to calculate mortality rates for specific queries, and then also the age adjustment process used in this report. The slight differences in the mortality rates between these two sources are due to rounding and the inclusion of deaths with unknown ages. Table 72 also shows the process of converting from crude mortality rates to age-adjusted mortality rates. While Maine has nearly double the crude rate of Utah, once the data are age adjusted, the rates become quite similar in all age cohorts.

Table 72: Comparison of Crude and Age-Adjusted Mortality Rates for Utah and Maine

<table>
<thead>
<tr>
<th>Age Cohort in Years</th>
<th>Standard Million Population</th>
<th>Utah (Younger Population State)</th>
<th>Maine (Older Population State)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Deaths</td>
<td>Population</td>
</tr>
<tr>
<td>Under 1</td>
<td>13,818</td>
<td>4,383</td>
<td>840,336</td>
</tr>
<tr>
<td>1–4</td>
<td>55,317</td>
<td>886</td>
<td>3,256,680</td>
</tr>
<tr>
<td>5–9</td>
<td>72,533</td>
<td>501</td>
<td>3,862,337</td>
</tr>
<tr>
<td>10–14</td>
<td>73,032</td>
<td>600</td>
<td>3,669,490</td>
</tr>
<tr>
<td>15–19</td>
<td>72,169</td>
<td>1,972</td>
<td>3,698,192</td>
</tr>
<tr>
<td>20–24</td>
<td>66,478</td>
<td>3,047</td>
<td>4,021,230</td>
</tr>
<tr>
<td>25–34</td>
<td>135,573</td>
<td>6,974</td>
<td>6,764,886</td>
</tr>
<tr>
<td>35–44</td>
<td>162,613</td>
<td>9,417</td>
<td>5,545,348</td>
</tr>
<tr>
<td>45–54</td>
<td>134,834</td>
<td>16,789</td>
<td>4,839,035</td>
</tr>
<tr>
<td>55–64</td>
<td>87,247</td>
<td>25,609</td>
<td>3,575,069</td>
</tr>
<tr>
<td>65–74</td>
<td>66,037</td>
<td>36,825</td>
<td>2,198,826</td>
</tr>
<tr>
<td>75–84</td>
<td>44,841</td>
<td>62,962</td>
<td>1,308,643</td>
</tr>
<tr>
<td>85 and over</td>
<td>15,508</td>
<td>72,608</td>
<td>475,018</td>
</tr>
<tr>
<td>All Years</td>
<td>1,000,000</td>
<td>242,573</td>
<td>44,055,090</td>
</tr>
</tbody>
</table>

Table 73: Comparison of Calculated Age-Adjusted Rates with CDC Wonder Reported Rates

<table>
<thead>
<tr>
<th>State</th>
<th>Utah</th>
<th>Maine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Type</td>
<td>Crude</td>
<td>Age-Adjusted</td>
</tr>
<tr>
<td>Calculated Rates</td>
<td>550.6</td>
<td>733.8</td>
</tr>
<tr>
<td>CDC Wonder</td>
<td>550.7</td>
<td>734.1</td>
</tr>
</tbody>
</table>
Data Suppression and Smoothing

Accommodations for Suppressed Values

The National Center for Health Statistics of the Centers for Disease Control and Prevention prohibits the reporting of death counts and death rates when the unit is fewer than ten cases. This results in rates that are suppressed for many counties in the Appalachian Region, as well as elsewhere in the United States. To account for this issue, data for counties with deaths below this threshold were augmented by incorporating data from nearby counties. To calculate the augmented rate for any particular county, a proportion of deaths (numerator) and population (denominator) from nearby counties were added to the base numbers for the target county. The proportion of a nearby county’s information that was added to the target’s decreases with increasing distance between the two.

This technique is called smoothing. In addition to overcoming suppression issues, it helps correct for unstable measures resulting from small population sizes. When small populations lead to even smaller numerators—such as occurs in counties with fewer than ten poisoning deaths in a given year—rates can be smoothed to eliminate statistical instability. The ten-death criteria for spatial adjustment was applied not only for the total number of deaths for any given mortality measure, but for individual age cohorts, as well.

Overview of Smoothing Process

The smoothing process applies weights to counties surrounding the county with the suppressed value, the “county of interest” in Figure 189 below. The process results in an augmented value resulting from the combination of actual data from the county of interest with weighted values from nearby counties. The degree to which a nearby county contributes to the augmented value depends on the distance from the county of interest to any given nearby county. Figure 189 below demonstrates this principle. The county of interest receives a weight of 1.0 and is fully shaded. Closer counties (e.g., the small, dark gray county directly to the southeast) have darker shading and higher weights than more distant counties (e.g., the large, white county further to the southeast). For each county, the number of deaths and populations in each age category then multiplies the weights. These numbers are then aggregated across counties by age category, and ultimately an age-adjusted rate for the county of interest is calculated.
Figure 189: Example of Weighting Counties for Smoothing Based on Distance

- County of Interest
- Lines represent distances between county centroids
- Shading represents weighting
Technical Detail

This approach is one we have used in the past (Ricketts & Holmes, 2007). First, we specified the general form of the weighting function:

\[
\text{WEIGHT} = \exp(\lambda \times \text{MILES})
\]

where \(\text{MILES}\) is the straight-line distance between two county centroids. The parameter \(\lambda\) is constrained to be negative, so the weighting function is equal to one if \(\text{MILES}\) equals zero, is decreasing in \(\text{MILES}\), and is bounded from below by zero. Values of \(\lambda\) that are closer to zero lead to a slower decay—for example, a \(\lambda\) of -0.05, will place greater weight on distant counties than a \(\lambda\) of, say, -0.2. In selecting the value of \(\lambda\), there is a tradeoff between faster decay (averaging over counties that are located closer to the county with the suppressed value) and using more counties (leading to more precise estimates). This approach significantly reduces the number of unreportable area-cause death rates while maximizing local influence on the augmented rate.

We conducted a grid search for the optimal \(\lambda\) using the following approach. First, we randomly chose counties in the Appalachian Region to be labeled as suppressed. We then calculated smoothed rates using the above methods for a variety of potential \(\lambda\). We repeated this exercise 1,000 times using a different \(\lambda\) each time. After each iteration, we calculated the mean squared error, or the average squared difference between the smoothed rate and the actual rate (the randomly chosen counties labeled as suppressed in this exercise had actual rates and served as the basis for comparison). This method allowed us to identify the \(\lambda\) with the smallest mean squared error. The \(\lambda\) satisfying this condition was \(-0.125\). We specified the same \(\lambda\) for all mortality rates.

With \(\lambda\) now in hand, the approach was as follows. For each suppressed county, we calculated the distance (\text{MILES}) between that county centroid and the county centroid of all other counties. We then calculated the weight associated with each county using the weighting function noted above.

Any county with a weight of less than 0.01 was dropped from the analysis. This approach has little practical effect on the augmented rate and eases any rounding issues (e.g., a populous county that is very distant may still have more weight than a nearby county of moderate size). We aggregated the deaths and population (numerator and denominator) across nearby counties—incorporating their weights—to create an augmented numerator and denominator for the suppressed county. This augmentation includes any deaths and population reported for the suppressed county, as well. At this point, the suppressed county has a “smoothed” number of deaths and population, with closer counties contributing more to the value. The mortality rate can then be calculated directly from these numbers.

For age-adjusted mortality rates (everything except infant mortality), the augmentation occurs prior to age-adjusting. Table 74 illustrates a simple example in which only two ages are considered: children and adults. Five counties are displayed, along with the number of deaths and population for both age groups. Distance from County A’s centroid determines the weight for each county. The final four columns calculate the weighted deaths and population. Note that County E, with a weight of less than 0.01, does not contribute to the aggregation.
Table 74: Sample County Illustration of Augmentation for Suppressed Data

<table>
<thead>
<tr>
<th>County</th>
<th>Deaths</th>
<th>Population</th>
<th>Age Group</th>
<th>Distance in Miles</th>
<th>Weight</th>
<th>Children Wtd deaths</th>
<th>Children Wtd pop</th>
<th>Adults Wtd deaths</th>
<th>Adults Wtd pop</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>1,000</td>
<td>Children</td>
<td>0</td>
<td>1.000</td>
<td>2.00</td>
<td>1000</td>
<td>18.00</td>
<td>10,000.0</td>
</tr>
<tr>
<td>A</td>
<td>18</td>
<td>10,000</td>
<td>Adults</td>
<td>0</td>
<td>1.000</td>
<td>18.00</td>
<td>10,000.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>1,200</td>
<td>Children</td>
<td>5</td>
<td>0.535</td>
<td>1.07</td>
<td>642.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>24</td>
<td>11,500</td>
<td>Adults</td>
<td>5</td>
<td>0.535</td>
<td>12.85</td>
<td>6,152.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>800</td>
<td>Children</td>
<td>10</td>
<td>0.287</td>
<td>1.15</td>
<td>229.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>37</td>
<td>7,900</td>
<td>Adults</td>
<td>10</td>
<td>0.287</td>
<td>10.60</td>
<td>2,267.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>1,500</td>
<td>Children</td>
<td>20</td>
<td>0.082</td>
<td>0.25</td>
<td>123.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>27</td>
<td>16,000</td>
<td>Adults</td>
<td>20</td>
<td>0.082</td>
<td>2.22</td>
<td>1,312.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>7</td>
<td>2,000</td>
<td>Children</td>
<td>40</td>
<td>0.007</td>
<td>weight &lt; .01 =&gt; 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>91</td>
<td>18,500</td>
<td>Adults</td>
<td>40</td>
<td>0.007</td>
<td>weight &lt; .01 =&gt; 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.47</td>
<td>1,994.6</td>
<td>43.7</td>
<td>19,731.8</td>
<td></td>
</tr>
</tbody>
</table>

After the aggregated deaths and population are calculated by age category, age-adjustment occurs using the approach outlined above (see Table 75). For purposes of this example, we specify weights of 0.1 (children) and 0.9 (adults) and thus calculate an age-adjusted rate of 221.7. Note that the example rates here—like the mortality rates in the report—are standardized per 100,000 population.

Table 75: Age-Adjusted Step for Spatial Adjustment

<table>
<thead>
<tr>
<th>Metric</th>
<th>Children</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality Rate</td>
<td>224.11</td>
<td>221.32</td>
</tr>
<tr>
<td>Age-adjusted weights (for example)</td>
<td>0.10</td>
<td>0.90</td>
</tr>
<tr>
<td>Weighted rate</td>
<td>22.41</td>
<td>199.32</td>
</tr>
</tbody>
</table>

**AUGMENTED AGE-ADJUSTED RATE** 221.73
Reciprocal Measures

In order to report health professional supply measures consistently, we calculated the reciprocal of values pulled directly from County Health Rankings. For example, 2016 County Health Rankings Data for Bibb County, Alabama shows a Primary Care Physician Ratio of 2,814:1 (persons per physician). We use the reciprocal (1/2,814), converted to primary care physicians per 100,000 people. The calculation is as follows:

\[
\frac{1}{2,814} \times 100,000 = .000355 \times 100,000 = 35.5
\]

Similarly, the dentist ratio of 5,627:1 becomes 17.8 dentists per 100,000. The only health professional supply measure that was not calculated in this manner was specialist physicians per 100,000, as County Health Rankings does not report this measure. Instead, the data for this measure come from the HRSA Area Health Resources File (AHRF).

Rurality in Appalachia

ARC, in coordination with staff at USDA’s Economic Research Service (ERS), has developed a simplified version of the 2013 Urban Influence Codes (UIC) for their research and analytical work. To create the UICs, the USDA starts with the official listing of metropolitan Core Based Statistical Areas (CBSAs/MSAs) published by the Office of Management and Budget (OMB). The USDA then distinguishes metropolitan counties by the population size of their metro areas, and nonmetropolitan counties by (i) the size of their largest city or town and (ii) proximity to metro areas. ARC simplified the USDA’s original 12-part county classification into the five levels described below.

1. **Large Metros**: Metropolitan counties in large metro areas of more than 1 million residents
2. **Small Metros**: Metropolitan counties in small metro areas of less than 1 million residents
3. **Non-metro, Adjacent to Large Metros**: Nonmetropolitan counties adjacent to a large metro area
4. **Non-metro, Adjacent to Small Metros**: Nonmetropolitan counties adjacent to a small metro area
5. **Rural (non-metro, not adj. to a metro)**: Nonmetropolitan counties not adjacent to a metro area

Table 76 illustrates the USDA twelve county classification system and the crosswalk from the USDA categories to the simplified five-class scheme.
Table 76: USDA Twelve County Classification System

<table>
<thead>
<tr>
<th>ARC Crosswalk Category</th>
<th>UIC Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metropolitan counties:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>In large metro area of 1+ million residents</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>In small metro area of less than 1 million residents</td>
</tr>
<tr>
<td><strong>Nonmetropolitan counties:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Micropolitan area adjacent to large metro area</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>Noncore adjacent to large metro area</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>Micropolitan area adjacent to small metro area</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>Noncore adjacent to small metro area and contains a town of at least 2,500 residents</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>Noncore adjacent to small metro area and does not contain a town of at least 2,500 residents</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>Micropolitan area not adjacent to a metro area</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>Noncore adjacent to micro area and contains a town of at least 2,500 residents</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>Noncore adjacent to micro area and does not contain a town of at least 2,500 residents</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>Noncore not adjacent to metro or micro area and contains a town of at least 2,500 residents</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>Noncore not adjacent to metro or micro area and does not contain a town of at least 2,500 residents</td>
</tr>
</tbody>
</table>

Population-weighted Averages for Large Geographies

To create measures for geographic units larger than the county, we used population-weighted averages. We used the same method of calculating these averages for all large geographies, including the United States, the Appalachian Region, the five Appalachian subregions, and then also the Appalachian, non-Appalachian, and total values for each state. The groupings of counties based on rurality and economic status also used this same methodology. These calculations are necessary because our data sources do not contain values for many of the report’s key geographies, such as the Appalachian Region or the Appalachian subregions. We thus needed a methodology to apply to all geographic groups in order to have a valid comparison. The population-weighted average approach provides a valid, consistent method.

The specific calculation used to generate population-weighted values is described here. Consider \( i \) counties that are members of an area \( A \). Area \( A \) could be a state, a specific economic status (e.g. distressed), or any other geography or grouping described in the report. The population-weighted average for variable \( x \) is calculated as the sum of the product of \( x \) and the population of each county, divided by the sum of the 2014 ACS populations of the counties in the area.
CREATING A CULTURE OF HEALTH IN APPALACHIA

B. Methodological and Technical Notes | APPENDICES

\[
AVERAGE_A = \left( \frac{1}{\sum_{i \in A} POP_i} \right) \sum_{i \in A} [POP_i \cdot x_i]
\]

As a result of the population weighting, values in this report for even commonly reported geographies—such as the United States as a whole—may differ slightly from published sources. However, using this population-weighted approach across all geographies ensures an accurate, consistent comparison.

**Time Periods of Mortality Measures**

For the mortality measures examined in this report, the values for each represent performance over a span of years: 2008–2014. This has the effect of reducing suppressed values in low population counties, as well as helping smooth single year spikes in mortality. These seven-year periods represent the original time period of the data as they are found in the CDC’s Compressed Mortality File.

**State-level effects of BRFSS Data**

Many of the variables derived from the CDC’s Behavioral Risk Factor Surveillance System (BRFSS) survey data show sharp state-level border effects. Because the survey is conducted at the state level, statistical modeling is utilized by County Health Rankings to develop these county-level values. Due to low numbers of responses in many counties, even with these statistical techniques, robust estimates for each county are not always possible. As such, county-level numbers oftentimes show little variation in any particular state. For more detailed information, please visit the County Health Rankings website.

**Box Plots**

A box plot is a type of graph that shows the distribution of data. Comparing box plots among different groups shows how the median of each group compares to the other groups, how much variation exists within each group, and how the variation compares between the groups.

**Figure 190: Components of a Box Plot**

"Dots" show unusually high or low values

"Whiskers" show the range of most values

The “box” stands for the middle half of all values

This line is the median: Half the values are above this amount, half are below.
The edges of the whiskers and the black line represent specific statistics calculated from the data. For example, the black line denotes the median (half of values are greater than this value, half are less than this value). The lower and upper edges of the box represent the 25th and 75th percentiles, respectively. The 25th percentile is the value for which 25 percent of county values are less, and the remainder (75 percent) are greater. The 75th percentile is defined similarly. The caps of the whiskers are defined as “adjacent values” (Tukey, 1977). The upper adjacent value (“top whisker”) is the largest observed value that is less than or equal to the 75th percentile plus 3/2 of the difference between the 75th and 25th percentile. The lower adjacent value is defined similarly. Outside values — the dots described as “unusually high or low values”—are those values that lie outside the adjacent values.