Gapminder: Investigating World Issues

Common Core Standards

**HSS.ID.B.6**
Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

**HSS.ID.B.6.B**
Informally assess the fit of a function by plotting and analyzing residuals.

**HSS.ID.B.6.C**
Fit a linear function for a scatter plot that suggests a linear association.

**HSS.ID.C.7**
Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

**HSS.ID.C.8**
Compute (using technology) and interpret the correlation coefficient of a linear fit.

**HSS.ID.C.9**
Distinguish between correlation and causation.

Materials
- At least one computer for each team of students
- Access to Excel or Statcrunch
- Task handout

Task Outline
- **Day 1**
  - Review positive, negative, and zero correlation. Review correlation and causation.
  - Discuss the issues we face in our world today
    - U.N. Millennium Development Goals
    - Mention how statistics can help us visualize these issues
  - Introduce Gapminder by watching the Hans Rosling video
    - Mention the impact of historical events on observable variables
    - Mention the importance of context in statistics
  - Demonstrate how to use Gapminder
  - Divide students into teams, hand out the worksheet, and explain the task
  - Have students begin the project
- **Day 2**
  - Review linear regressions and lines of best fit
  - Demonstrate how to copy/paste data into Excel or Statcrunch and fit a linear regression to the data
  - Have students continue the project
- **Days 3-5**
  - Have students work on their presentations
- **Day 6**
  - Presentations
Implementation
Note: This task is designed to be completed after a unit on correlation, causation, and linear regression.

Day 1
On the first day, start by reviewing types of correlation and correlation vs. causation. Students will likely work on these issues in the task today.

Introduce the subject by asking students about the problems they think we face in the world today. Mention that the U.N. developed a set of Millennium Development Goals referring to those problems, and ask students how they think statistics can help us understand the problems in the world.

Introduce Gapminder as a way to visualize data about the world. Watch the Hans Rosling video that demonstrates the main wealth vs. health graph on Gapminder World. Emphasize that historical events have an observable impact on measurable variables, and that statisticians often draw on outside information to develop research questions and predict outcomes.

Divide students into teams. Hand out the worksheets and Gapminder tutorials. Take some time to show students how Gapminder works. Some possible graphs and events to show are:

- Wealth vs. Health looks at life expectancy vs. income. Each bubble represents a country. The size of each bubble reflects the population of that country. Point out the United States, China, and India. Point out how the colors represent regions of the world. Ask students what they notice about the relationship between income and life expectancy, and why they think that is the case. Now, introduce the dimension of time. Possible events to point out:
  - 1918: The drop in life expectancy during WWI and the Spanish Influenza
  - 1933: The life expectancy in Russia drops due to famine
  - 1960: Famine following the Communist Revolution in China
  - 1970s: Cambodian Civil War
  - How many have heard the phrase "Africa is not a country"? Point out the huge regional variation in African nations. Mauritius has a higher per capita income than China and a life expectancy on par with China. Click on Botswana, which is a Democratic African nation. Track the progress of Botswana and point out the huge drop in life expectancy due to AIDS.
  - The outsourcing of labor in the U.S.

- The outsourcing of labor in the U.S.
- Graph Industry Workers (under the Work – Employment by Sector heading) over time for the U.S., China, and Vietnam
- Ask students to interpret the graph

Explain that there is a continuum of levels of development across countries in the world. Mention that a lot of organizations donate time and money to improve people’s quality of life in other countries. Explain the task: Students will develop a list of how to determine whether a country needs aid, select one country to receive aid, and present their results. There is no one right answer, but students must back up their results with data.

Have students begin by considering how they would measure “need,” what variables they might consider, and what relationships might be present between those variables. Each team should work on completing up to question 4 of their worksheet.

Ask students to share their answers to the first four questions.

**Day 2**
On the second day, begin by reviewing linear regression topics, including the idea of correlation versus causation. Students will consider correlation coefficients, linear regression models, and observation errors today.

Demonstrate how to download data from Gapminder into a statistical package such as Excel or Statcrunch. Demonstrate how to perform a linear regression in the software. Mention possible issues with missing values in data, and ask that students consider whether they have enough data to answer their questions.

Have students continue to work on the worksheet questions. Students should reach question 7 today.

**Days 3-5**
For the next few classes, have students continue to work on their linear regression models and their presentations.

**Day 6**
On the final day, have students present their results. Have each team critique the other teams’ work, and encourage students to ask questions.
**Selected Menu of Quantified Statistics**

*(Available for comparison at http://www.gapminder.org/world.)*

<table>
<thead>
<tr>
<th>Heading Categories</th>
<th>Society</th>
<th>Education</th>
<th>Economy</th>
<th>Infrastructure</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children Per Woman</td>
<td>Murder Rate</td>
<td>Spending Per Student</td>
<td>Poverty Rate</td>
<td>Broadband Subscribers</td>
<td>Population</td>
</tr>
<tr>
<td>CO2 Emissions</td>
<td></td>
<td>Literacy Rate</td>
<td>Aid Given</td>
<td>Cell Phones</td>
<td>Growth</td>
</tr>
<tr>
<td>Income per Person</td>
<td></td>
<td>8th Grade Math Achievement</td>
<td>Aid Received</td>
<td>Internet Users</td>
<td>Density</td>
</tr>
<tr>
<td>Child Mortality</td>
<td></td>
<td></td>
<td>Exports</td>
<td>Personal Computers</td>
<td>Median Age</td>
</tr>
<tr>
<td>Life Expectancy</td>
<td></td>
<td></td>
<td>Imports</td>
<td>% of Roads Paved</td>
<td></td>
</tr>
<tr>
<td><strong>Economy</strong></td>
<td></td>
<td></td>
<td>Agriculture as % of GDP</td>
<td>Traffic Deaths</td>
<td></td>
</tr>
<tr>
<td>Poverty Rate</td>
<td></td>
<td></td>
<td>Industry as % of GDP</td>
<td># of Cars, Trucks, Buses</td>
<td></td>
</tr>
<tr>
<td>Aid Given</td>
<td></td>
<td></td>
<td>Services as % of GDP</td>
<td>Traffic</td>
<td></td>
</tr>
<tr>
<td>Aid Received</td>
<td></td>
<td></td>
<td></td>
<td>Access to Clean Water</td>
<td></td>
</tr>
<tr>
<td>Exports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sources of Images:
http://www.globalpartnership.org/blog/kicking-500-day-rally-education
Create a Presentation: Investigating World Issues

Many national, international, and private organizations donated money to other nations in need to encourage development. In 2012, USAID gave over $48,400,000,000 in economic and military aid to other nations in the world. If you controlled the distribution of aid and could prioritize one country to aid, which country would you choose based on need?

Overall Goal:
In teams, create a presentation about which country you would recommend for aid, what issues you would focus on funding in the country, how you would distribute aid, and why you hope your contributions to these issues would have a positive impact on the country.

Discuss and Answer the Following Questions:
1. Using the U.N. Millennium Development Goals as a starting point, what larger issues (e.g., health, poverty, etc.) would you consider using to measure a country’s need for aid?
   
   a. List the larger issues you would use to judge how much a country needs aid. What variable(s) would you look at to measure those issues?

   b. Why did you choose to look at those variables? Look at the definition of each variable. What does each variable measure? Are there any dimensions of each issue that the variable(s) failed to capture?
2. Think about the possible relationships between the variables you’ve chosen. Pick two variables and predict how they are related (positively, negatively, etc.). As a check of your understanding:

   a. Sketch your prediction of how those two variables are related.

   b. Why did you predict that the variables would be related in that way?

3. Open up Gapminder, graph the two variables you selected in question 3, and compare the actual results to your predicted results. How did the two results compare? Did anything surprise you?

4. Have each member of your team ask at least one question he/she thinks would tell us more about which country is most in-need of aid based on the variables and issues you chose. List those questions here.
5. Open Gapminder. Explore the questions your team came up with.
   
a. Write a sentence describing what happened for at least three questions.

   b. Develop and list a set of criteria for how you will classify a country as “in-need of aid.”

6. Write down the country you would recommend to receive aid.

7. Choose two variables to investigate further (these may be based on prior questions your team asked, based on the variables you investigated in question 3, or new variables altogether). Download data for those two variables from the following link: [http://www.gapminder.org/data/](http://www.gapminder.org/data/)
   
a. List the two variables you chose:

   b. Choose a year to investigate and copy the data for those two variables across all countries in that year. Paste the data into StatCrunch or Excel. Make sure there is adequate data for countries in that year so that you can investigate the relationship between the two variables. Make the country you chose to receive aid is included in the dataset.

   c. Fit a linear regression to the data to show the relationship between the two variables based on the countries of the world. Save the result (graph, model formula, and correlation coefficient). Do you think that a linear regression model adequately describes the relationship between the two variables? Why or why not?
d. As a check of your understanding, discuss the issue of proving a causal relationship between the two variables you investigated.

8. Based on the issues you chose to measure countries’ need, the variables you selected to tell us about those issues, and the questions you asked, create a presentation backed by data that tells us which country you think is most in need of aid.

In your presentation, please:

a. Use two outside sources to find information about the history and government of the country you looked at and tie that information into your presentation.

b. Discuss the investigations you conducted and the criteria you used (issues, variables, measurements on those variables) to determine whether a country needed aid. Discuss why you selected those criteria and whether your criteria matched the U.N. Millennium Development Goals. Use data to answer:

   i. How did the issues you looked at affect the world?
   ii. How did the issues you looked at affect the country you chose? Did what happened in your chosen country match what you expected to see based on how those issues affected the world? Why do you think that was the case?

c. Show the more in-depth linear regression results from question 6.

   i. Discuss the general implications of the relationship between the two variables in the world.
   ii. Locate the data point representing your chosen country.
   iii. Discuss how your country does or does not fit the implications of the model. Point out the difference between what we’d expect to see based on the linear model and what we actually see based on the data for your country (this is the observation’s error).

d. Think about the logistics of how you would distribute aid to the country (would you entrust money to its government, send supplies, hire personnel, etc.?)) and be prepared to discuss these issues.