

Cancer: A Global View

Key Facts¹

- Cancer is a leading cause of death worldwide: it accounted for 7.4 million deaths (around 13% of all deaths) in 2004.
- Lung, stomach, colorectal, liver, and breast cancer cause the most cancer deaths each year.
- The most frequent types of cancer differ between men and women.
- More than 30% of cancer deaths can be prevented.¹
- Tobacco use is the single most important risk factor for cancer.
- Cancer arises from a change in one single cell. The change may be started by external agents and inherited genetic factors.

Learning Objectives

After completion of this module, the student will be able to

1. use Gapminder to explore “social, economic and environmental development at local, national and global levels”
2. use WHOSIS to download global health data
3. perform and interpret logarithmic transformations for graphical display
4. make predictions based on a fitted trendline

Knowledge and Skills

1. logarithmic transformation
2. continuous time population models
3. fitting a straight line to data

Prerequisites

1. calculating percent changes
2. straight lines
3. natural logarithm, exponential function
4. graphing in EXCEL
5. fitting a straight line to data points in EXCEL and displaying the equation

¹ Source: World Health Organization <http://www.who.int/mediacentre/factsheets/fs297/en/>

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:**

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was partially supported by a HHMI Professors grant from the Howard Hughes Medical Institute. Page 1

Gapminder

Gapminder (<http://www.gapminder.org/>) “is a non-profit venture promoting sustainable global development and achievement of the United Nations Millennium Development Goals by increased use and understanding of statistics and other information about social, economic and environmental development at local, national and global levels.” (<http://www.gapminder.org/about-gapminder/>)

In-class Activity 1

Watch the following video in Gapminder: <http://www.gapminder.org/videos/lung-cancer-statistics/>

In-class Activity 2

Go to Gapminder (<http://www.gapminder.org/>). To get a sense for how various variables depend on population size, explore **Gapminder World**: Display *Income per capita (GDP/ capita, inflation-adjusted \$)* on the horizontal axis and the following secondary variables on the vertical axis.

- *Health > Cancer, lung > Lung cancer, new cases per 100,000 men*
- *Health > Cancer, lung > Lung cancer, new cases per 100,000 women*
- *Health > Cancer, liver > Liver cancer, new cases per 100,000 men*
- *Health > Cancer, liver > Liver cancer, new cases per 100,000 women*
- *Health > Cancer, colon&rectum > Colon&Rectum, new cases per 100,000 men*
- *Health > Cancer, colon&rectum > Colon&Rectum, new cases per 100,000 women*
- *Births and deaths > Population growth (annual %)*
- *Children per woman (total fertility)*

Circles of different colors and sizes are shown in the graph. Explain what the size and the color of a circle means. Use the **Play** button to see how each of the variables changes over time. Hold your cursor over geographic regions. Explore what happens when you change the axes from **lin** to **log**.

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:**

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was partially supported by a HHMI Professors grant from the Howard Hughes Medical Institute. Page 2

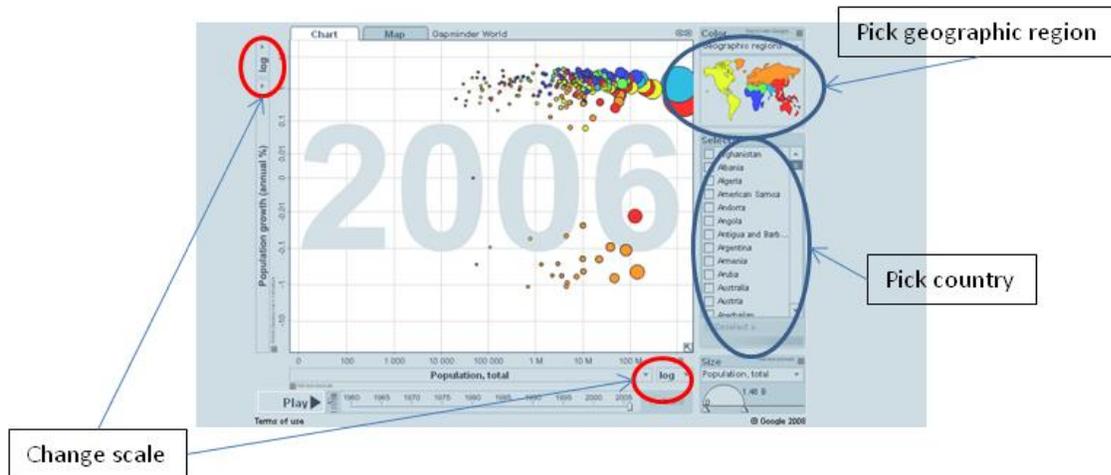


Figure 1: Screenshot of Gapminder to explore Gapminder

In-class Activity 3

Explore whether there are any relationships among the indicators you investigated in In-class Activity 2. Feel free to add other indicators.

To download data from Gapminder, go to List of indicators and download the “indicator list” to excel. The 4th tab (Indicators) lists all indicators for which Gapminder has data. To download a specific indicator, click on the appropriate cell in Column E (Indicator url). You can then download the data to a spreadsheet.

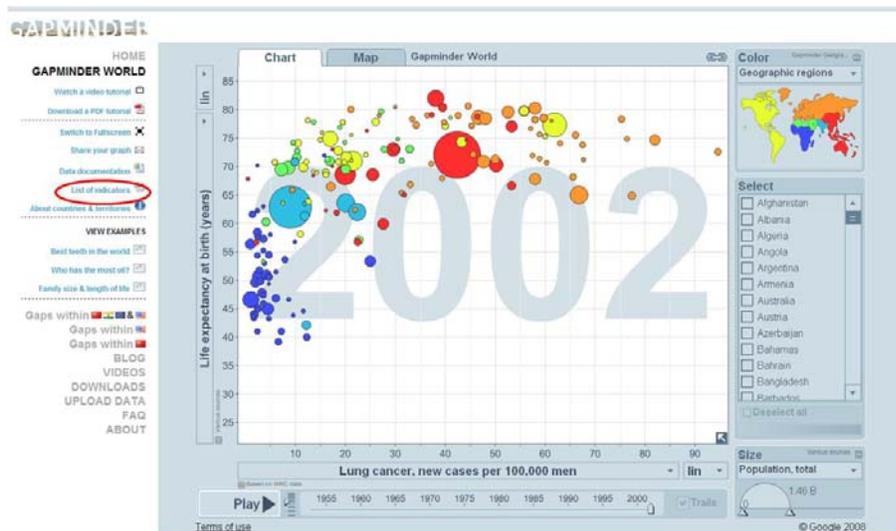


Figure 2: Screenshot of Gapminder to download data

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:**

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was partially supported by a HHMI Professors grant from the Howard Hughes Medical Institute. Page 3

Transformations

Campbell (1996), page 1109, shows a graph that illustrates the decrease in fecundity at high population densities in a small herb, plantain (*Plantago major*). The data can be found in the spreadsheet under the Plantain tab. On the horizontal axis is the number of seeds planted per m², on the vertical axis is the average number of seeds per reproducing individual (Figure 3). We make two observations: (1) The axes are on a scale where multiples of 10 are equidistant. (2) The points seem to follow a straight line.

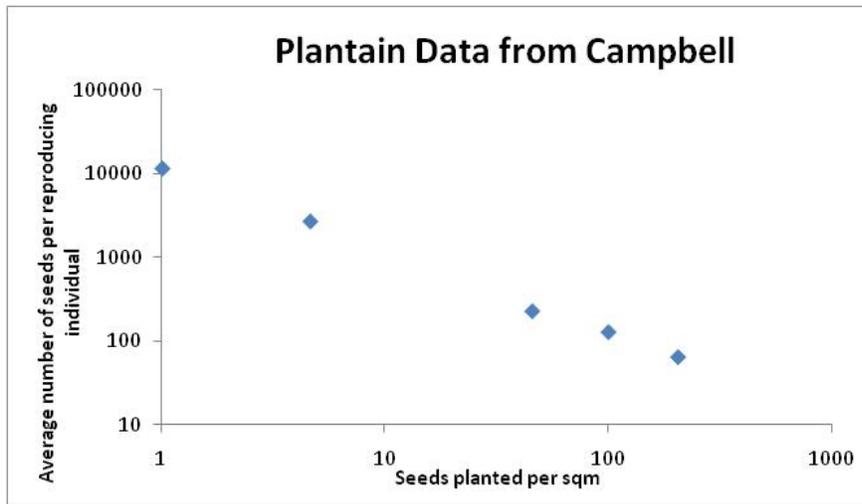


Figure 3: Data points on a graph where both axes are logarithmically transformed. (Drawn after Campbell, 1996².)

On either axis, the numbers span several orders of magnitude. Graphing the points in this way makes it easier to see the relationship since the data almost follow a straight line on this graph.

Data from the Christmas Bird Count was used in Van Bael and Pruett-Jones³ as a proxy for population size. The data can be found in the spreadsheet under the Parakeet tab. If we use a logarithmic scale on the vertical axis, then a straight line fits the data points. The graph is shown in Figure 4.

² Campbell, N.A. (1996) *Biology*. Fourth Edition. The Benjamin/Cummings Publishing Company, Inc.

³ Van Bael, S. and S. Pruett-Jones. 1996. Exponential population growth on Monk Parakeets in the United States. *Wilson Bulletin* 108(3):584-588.

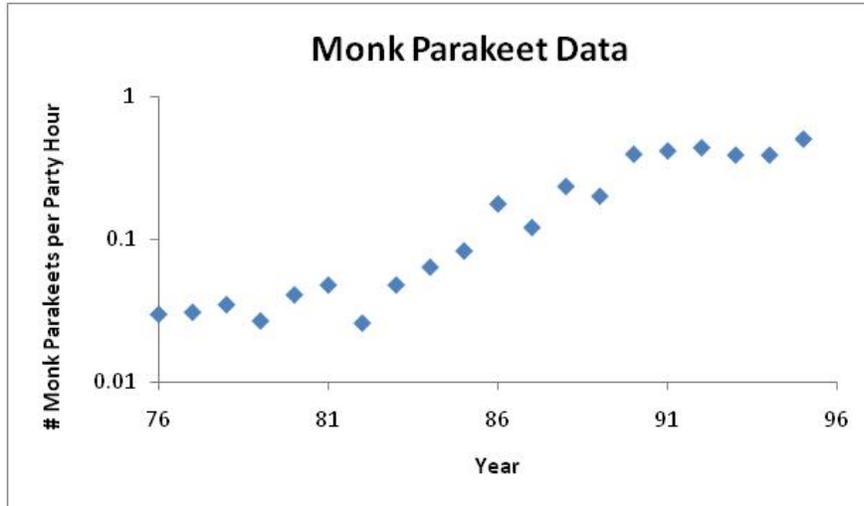
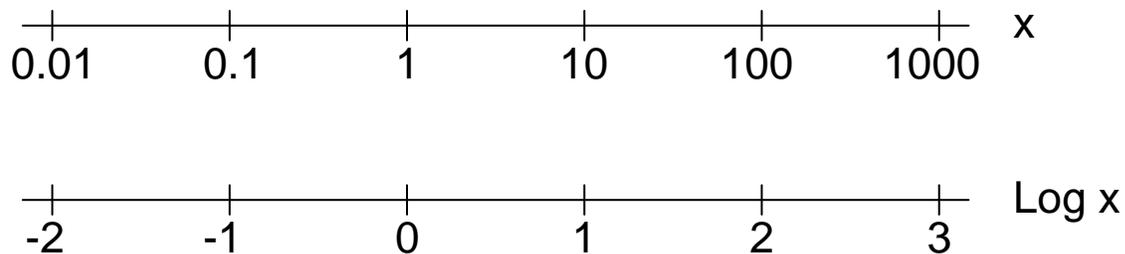


Figure 4: Data points on a graph where the vertical axis is logarithmically transformed.

We will learn what functions are used to describe linear relationships after the data is transformed logarithmically. The first step in understanding these relationships is to gain familiarity with logarithms to base 10.

The Logarithmic Scale

A scale where multiples of 10 are equidistant as in the graph above is called a *logarithmic scale*. It is called logarithmic since the logarithms of the labels on the axis below are equidistant (here: $\log x$ or $\text{Log } x$ means the logarithm to base 10):



In-class Activity 4

(a) On the two axes above find the following numbers: $x=0.05, 0.2, 8, 15, 750$. (b) Why do you think we choose logarithms to base 10, instead of some other base? (c) Can you plot negative numbers on a logarithmic scale? (d) As x approaches 0, where would you find x on a logarithmic scale?

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:**

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was partially supported by a HHMI Professors grant from the Howard Hughes Medical Institute. Page 5

The two most frequent transformations of a relationship $y = f(x)$ are (1) both axes are logarithmically transformed or (2) the y -axis is logarithmically transformed and the x -axis is on an arithmetic (linear) scale. In either case, when such a transformation results in a straight line, we can find the analytical form of the relationship.

Case 1: Both axes are logarithmically transformed

If the relationship between $\log y$ and $\log x$ is linear, we can write

$$\log y = B + a \log x$$

where B is the intercept on the vertical axis and a is the slope.

$$\log y = B + a \log x$$

$$y = 10^{[B+a \log x]}$$

$$y = 10^B \underbrace{10^{a \log x}}_{10^{\log x^a} = x^a}$$

With $b = 10^B$, we can now write

$$y = bx^a$$

This is a power function. We can summarize this result.

If both axes are logarithmically transformed and a straight line results, then the relationship between x and y is a **power function**: $y = bx^a$

Case 2: The x -axis is on an arithmetic scale and the y -axis is logarithmically transformed

If the relationship between $\log y$ and x is linear, we can write

$$\log y = C + mx$$

where C is the intercept on the vertical axis and m is the slope.

$$\log y = C + mx$$

$$y = 10^{[C+mx]}$$

$$y = 10^C \underbrace{10^{mx}}_{(10^m)^x}$$

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:**

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was partially supported by a HHMI Professors grant from the Howard Hughes Medical Institute. Page 6

With $c = 10^C$ and $a = 10^m$, we can now write

$$y = ca^x$$

This is an exponential function. We can summarize this result.

If the x -axis is on an arithmetic scale and the y -axis is logarithmically transformed and a straight line results, then the relationship between x and y is an **exponential function**:

$$y = ca^x$$

In-class Activity 5a

Graph each of the following functions in Excel and transform the axes so that a straight line results. Use the Trendline option to fit a straight line. Display the equation on the chart.

(a) $y = 2x^3$

(b) $y = 2.5 \times 3^x$

In-class Activity 5b

Fit functions to both the Plantain and Parakeet data in the spreadsheet.

World Health Organization

“WHOSIS, the WHO Statistical Information System, is an interactive database bringing together core health statistics for the 193 WHO Member States. It comprises more than 100 indicators, which can be accessed by way of a quick search, by major categories, or through user-defined tables. The data can be further filtered, tabulated, charted and downloaded. The data are also published annually in the World Health Statistics Report released in May.” (<http://www.who.int/whosis/en/>)

Click on “Select from a list of regions/countries, indicators and time periods” and select

- Countries: Select all countries
- Indicators: Select Age-standardized mortality rate for cardiovascular diseases (per 100 000 population) and Age-standardized mortality rate for cardiovascular diseases (per 100 000 population)
- Time periods: Select all time periods

Click on “Create Table.” WHOSIS creates a table for the year 2002, the only year where data for cancer and cardiovascular disease is available. To analyze the data, export the data to a .csv file (see Figure 5)

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:**

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was partially supported by a HHMI Professors grant from the Howard Hughes Medical Institute. Page 7

The screenshot shows the WHO WHOSIS database search interface. The page title is "World Health Organization" and "WHOSIS WHO Statistical Information System". The search results are displayed in a table format. The table has three columns: "Location", "Age-standardized mortality rate for cancer (per 100 000 population)", and "Age-standardized mortality rate for cardiovascular disease (per 100 000 population)". The data is for the year 2002. The "Export (.csv)" button is circled in red.

Location	Age-standardized mortality rate for cancer (per 100 000 population)	Age-standardized mortality rate for cardiovascular disease (per 100 000 population)
Algeria	103.0	314.0
Angola	179.0	486.0
Benin	154.0	432.0
Botswana	124.0	339.0
Burkina Faso	162.0	459.0
Burundi	146.0	439.0
Cameroon	150.0	436.0
Cape Verde	127.0	356.0
Central African Republic	154.0	445.0
Chad	156.0	443.0
Comoros	128.0	381.0
Congo	134.0	393.0
Cote d'Ivoire	160.0	436.0
Democratic Republic of the Congo	161.0	465.0

Figure 5: Screenshot of database search. Click on Export (circled in red) to export to .csv file.

Mashups

A mashup is an “application that combines data or functionality from two or more external sources to create a new service.” (Source: <http://www.programmableweb.com/fag>). Mashups are a great way to engage students in developing their own learning environment. Instead of providing students with a single textbook and handouts, students are asked to combine materials from different sources to drive their own inquiry.

Tobacco has been implicated as a major risk factor for lung cancer. The Center for Disease Control and Prevention states that “in the United States, about 90% of lung cancer deaths in men and almost 80% of lung cancer deaths in women are due to smoking. People who smoke are 10 to 20 times more likely to get lung cancer or die from lung cancer than people who do not smoke.”⁴ Up to 40% of newly diagnosed lung cancers occur in former smokers (median abstinence duration 9 year)⁵.

⁴ Source: http://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm

⁵ Source: <http://www.chestx-ray.com/Smoke/Smoke.html>

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:**

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was partially supported by a HHMI Professors grant from the Howard Hughes Medical Institute. Page 8

In-class Activity 6

We can ask the question whether the price of tobacco products affect lung cancer rates. A mashup of data from the CDC, World Health Organization (WHO), and Gapminder provides data for select countries on the following three indicators: (1) price of 100 packs of cigarettes as a percentage of GDP per capita, (2) the per capita GDP for select years, and (3) lung cancer rates in men and women for a single year. (See MashupCigarettesGDPCancer.xls.) Use the data to investigate the relationship between the three indicators.

In-class Activity 7

The data set in In-class Activity 6 is missing data from the United States. (a) Find the U.S. per capita GDP for a recent year and the price of cigarettes. (b) Calculate the Price of 100 packs of cigarettes as a percentage of GDP per capita and predict the number new lung cancer cases per 100,000 male population. (c) Based on the trendlines you found in In-class Activity 6, how much would a pack of cigarettes need to cost to reduce new lung cancer cases to about half of what they currently are? (d) Critique the approach in (c).

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:**

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was partially supported by a HHMI Professors grant from the Howard Hughes Medical Institute. Page 9