

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.



Figure 1: Correlation between smoking and lung cancer in US males, showing a 20-year time lag between increased smoking rates and increased incidence of lung cancer.

(http://en.wikipedia.org/wiki/File:Cancer_smoking_lung_cancer_correlation_from_NIH.svg)

Learning Objectives

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

1. explore “social, economic and environmental development at local, national and global levels” with Gapminder
2. perform and interpret logarithmic transformations for graphical display
3. download global health data from Gapminder and WHO

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:** February 28, 2010; June 7, 2010

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License, which permits unrestricted use, distribution, and reproduction in any medium, and allows others to translate, make remixes, and produce new stories based on this work, provided the original author and source are credited and the new work will carry the same license.

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 cancer rates depend on per capita income explore **Gapminder World**: Display *Income per capita (GDP/ capita, inflation-adjusted \$)* on the horizontal axis and the following secondary variables on the vertical axis (see Figure 2).

- *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*

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**.

Do you detect any patterns? Discuss in your group.

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:** February 28, 2010; June 7, 2010

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License, which permits unrestricted use, distribution, and reproduction in any medium, and allows others to translate, make remixes, and produce new stories based on this work, provided the original author and source are credited and the new work will carry the same license.

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

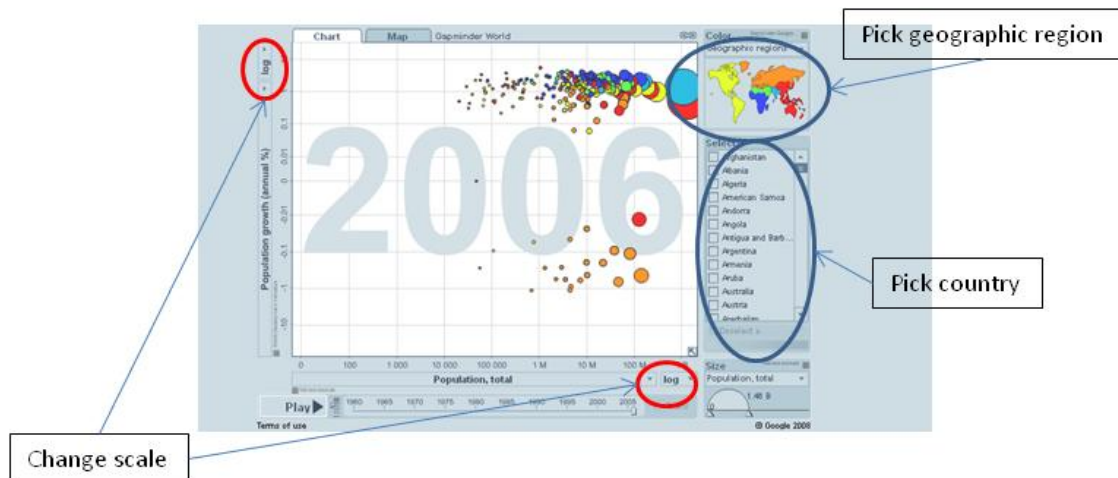


Figure 2: Screenshot of Gapminder to explore Gapminder

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.

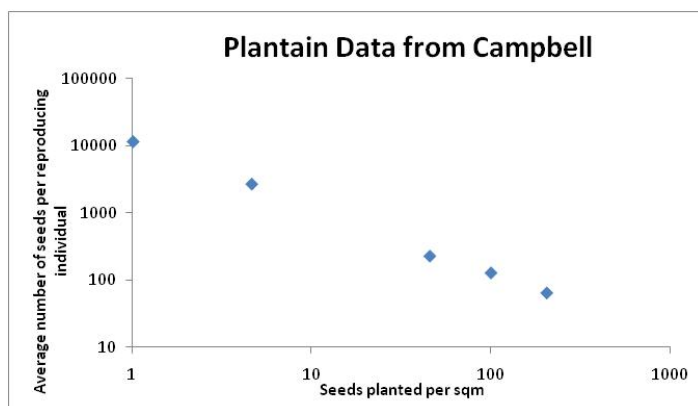


Figure 3: Data points on a graph where both axes are logarithmically transformed. (Drawn after Campbell, 1996².)

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

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:** February 28, 2010; June 7, 2010

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License, which permits unrestricted use, distribution, and reproduction in any medium, and allows others to translate, make remixes, and produce new stories based on this work, provided the original author and source are credited and the new work will carry the same license.

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

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.

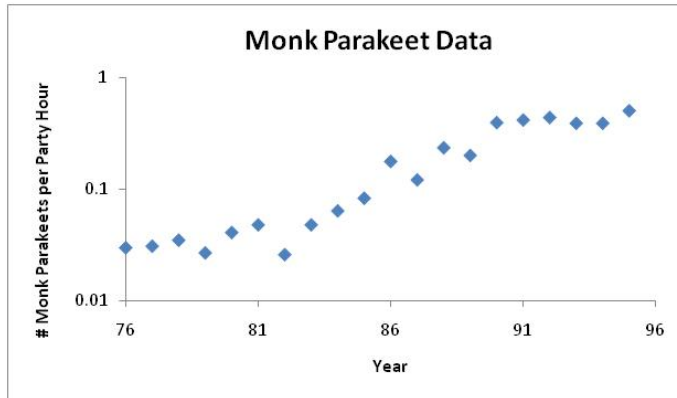
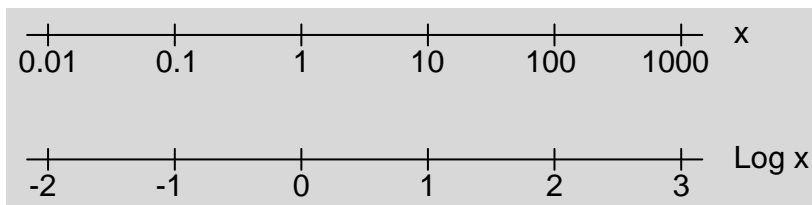


Figure 4: Data points on a graph where the vertical axis is logarithmically transformed.

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):



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.

Citation: Neuhauser, C. *Cancer Global View*.

Created: June 19, 2009 **Revisions:** February 28, 2010; June 7, 2010

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License, which permits unrestricted use, distribution, and reproduction in any medium, and allows others to translate, make remixes, and produce new stories based on this work, provided the original author and source are credited and the new work will carry the same license.

Funding: This work was partially supported by a HHMI Professors grant from the Howard Hughes Medical Institute. Page 4

In-class Activity 3

(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?

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

$$\log y = C + mx$$

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:** February 28, 2010; June 7, 2010

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License, which permits unrestricted use, distribution, and reproduction in any medium, and allows others to translate, make remixes, and produce new stories based on this work, provided the original author and source are credited and the new work will carry the same license.

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

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

$$\log y = C + mx$$

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

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

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 4

- (a) Fit an appropriate function to the Plantain data in the spreadsheet.
- (b) Fit an appropriate function to the Parakeet data in the spreadsheet.

Downloading Data

Gapminder

To download data from Gapminder, click on the DATA tab in Gapminder World. You can then download the data to a spreadsheet. You can also view data directly by clicking on “View Data” icon below each indicator (see Figure 5).

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:** February 28, 2010; June 7, 2010

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License, which permits unrestricted use, distribution, and reproduction in any medium, and allows others to translate, make remixes, and produce new stories based on this work, provided the original author and source are credited and the new work will carry the same license.

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

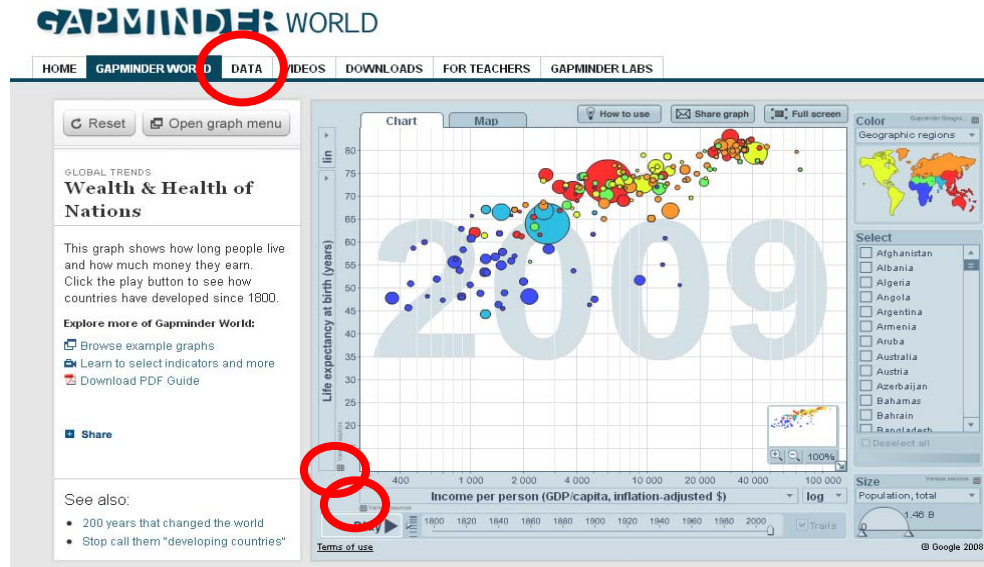


Figure 5: Screenshot of Gapminder to download data

World Health Organization

“The Global Health Observatory (GHO) is WHO's portal providing access to data and analyses for monitoring the global health situation. It provides critical data and analyses for key health themes, as well as direct access to the full database. The GHO presents data from all WHO programmes and provides links to supporting information.” (<http://www.who.int/gho/en/>)

Click on “Database” in the left-hand column and “Browse the GHO database”

- 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 6).

Citation: Neuhauser, C. Cancer Global View.

Created: June 19, 2009 **Revisions:** February 28, 2010; June 7, 2010

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License, which permits unrestricted use, distribution, and reproduction in any medium, and allows others to translate, make remixes, and produce new stories based on this work, provided the original author and source are credited and the new work will carry the same license.

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

| Location | Time Period | Life expectancy at birth (years) ↑ | Healthy life expectancy (HALE) at birth (years) ↑ |
|---------------------|-------------|------------------------------------|---|
| Afghanistan | 2007 | | 42 |
| Albania | 2007 | | 72 |
| Algeria | 2007 | | 71 |
| Andorra | 2007 | | 81 |
| Angola | 2007 | | 53 |
| Antigua and Barbuda | 2007 | | 74 |
| Argentina | 2007 | | 75 |
| Armenia | 2007 | | 69 |
| Australia | 2007 | | 82 |
| Austria | 2007 | | 80 |
| Azerbaijan | 2007 | | 68 |
| Bahamas | 2007 | | 74 |
| Bahrain | 2007 | | 75 |
| Bangladesh | 2007 | | 64 |
| Barbados | 2007 | | 76 |
| ... | --- | | -- |

Figure 6: 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/faq>). 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 or 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.

Group Project: Summative Assessment

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.

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:** February 28, 2010; June 7, 2010

Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License, which permits unrestricted use, distribution, and reproduction in any medium, and allows others to translate, make remixes, and produce new stories based on this work, provided the original author and source are credited and the new work will carry the same license.

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