Applied Epidemiology I: Graphs

Enoch Yi-Tung Chen

Department of Medical Epidemiology and Biostatistics, Karolinska Insitutet

December 9, 2020

Acknowledgements

This course material is based on my learning from Anastasia Lam's teachings in last year's Applied Epidemiology I lab sessions, and readings from *Epidemiology* by Gordis [1], *A First Course in Probability and Statistics* by Goldsman and Goldsman [2], *Principles of Biostatistics* by Pagano and Gauvreau [3], and *Biostatistics I* by Gabriel and Frumento [4]. I especially want to thank Marlene Stratmann for reviewing the slides and Prof. Paul Dickman for providing me with suggestions to improving the teaching.

Outline

Graphs: Bad examples

Graphs can say more than texts! But it depends.....
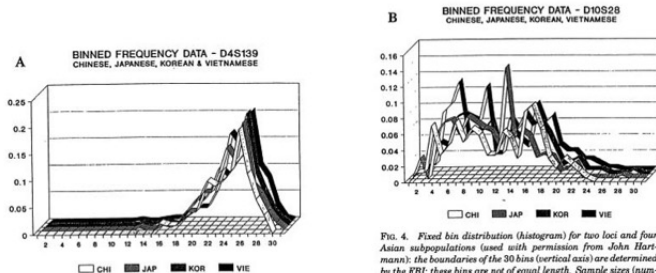Sometimes less is more.

Fig. 4. *Fixed bin distribution (histogram) for two loci and four Asian subpopulations (used with permission from John Hartmann): the boundaries of the 30 bins (vertical axis) are determined by the FBI; these bins are not of equal length. Sample sizes (numbers of individuals) for Chinese, Japanese, Korean and Vietnamese are 103, 125, 93 and 215 for D4S139 and 120, 137, 100 and 193 for D10S28. The horizontal axis is the bin number; bins are not of equal length.*
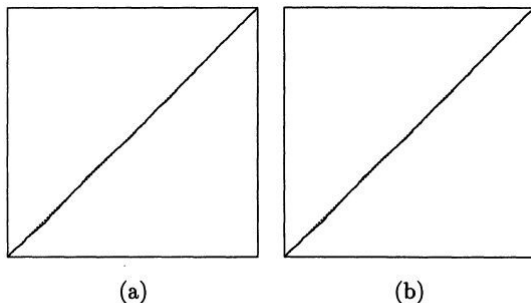
Graphs: Bad examples

Too fancy?
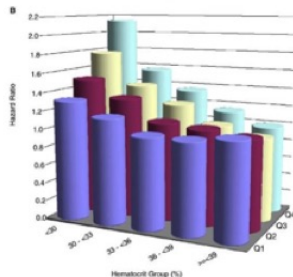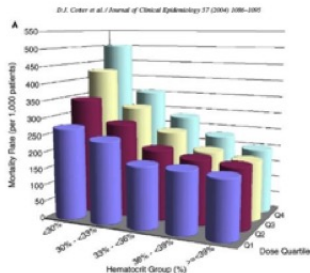
Insufficient info?

(a)   (b)

Figure 1. SRQ Plots of T_i/T_n (Vertical Axes) Against i/n (Horizontal Axes) for the Gibbs Sampler (a) and an Alternating Gibbs/Independence Sampler (b) for the Pump Failure Data Based on Runs of Length 5,000. Lines through the origin with unit slope are shown dashed; axis ranges are from 0 to 1 for all axes.

Graphs: Bad examples

Sometimes there is no right nor wrong, it just depends on your interest.
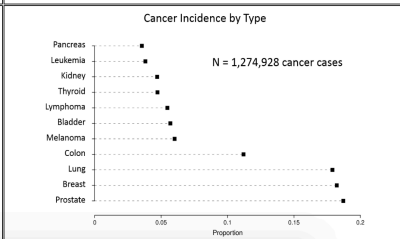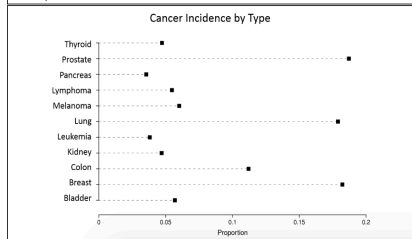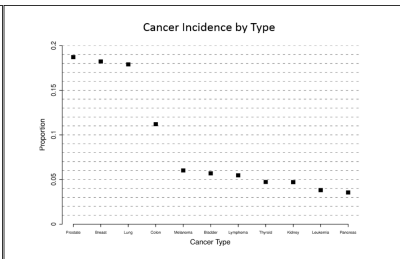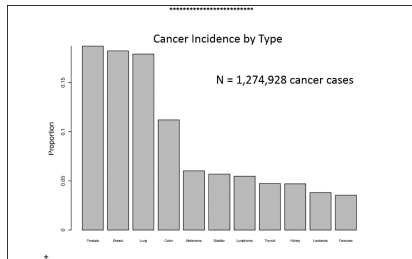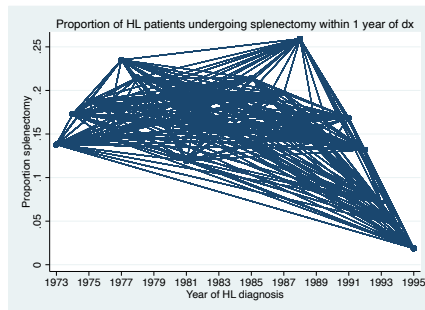
Graphs: Learning from errors

Which part went wrong here?
Hint: something was missed in the code.
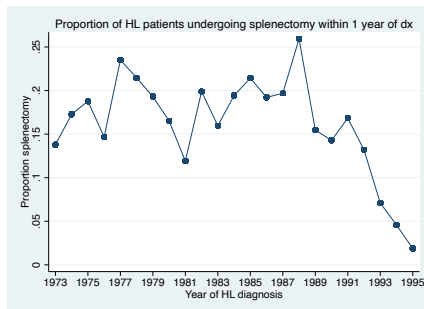
```
twoway connected prop diagyear, ///
subtitle("Proportion of HL patients") ///
ytitle(Proportion splenectomy) ///
xlabel(1973(2)1995)
```

Graphs: Learning from errors

It makes such a big difference if you missed `sort`!

```
twoway connected prop diagyear, ///
subtitle("Proportion of HL patients") ///
ytitle(Proportion splenectomy) ///
xlabel(1973(2)1995) ///
sort
```

Graphs: Basics of making graphs

- Headings should be
    - self-explanatory and informative
    - placed below the graphs.
    - (Table headings are placed above.)

Graphs: Basics of making graphs

- Headings should be
    - self-explanatory and informative
    - placed below the graphs.
    - (Table headings are placed above.)
- Graphs are numbered sequentially.

Graphs: Basics of making graphs

- Headings should be
  - self-explanatory and informative
  - placed below the graphs.
  - (Table headings are placed above.)
- Graphs are numbered sequentially.
- Allowed to copy figures from the output of software.

Graphs: Basics of making graphs

- Headings should be
    - self-explanatory and informative
    - placed below the graphs.
    - (Table headings are placed above.)
- Graphs are numbered sequentially.
- Allowed to copy figures from the output of software.

- Grey scales + differentiable shapes
- Colors + differentiable shapes

Graphs: Basics of making graphs

- Headings should be
    - self-explanatory and informative
    - placed below the graphs.
    - (Table headings are placed above.)
- Graphs are numbered sequentially.
- Allowed to copy figures from the output of software.

- Grey scales + differentiable shapes
- Colors + differentiable shapes
- x-axis and y-axis are clearly defined with units.

- Headings should be
  - self-explanatory and informative
  - placed below the graphs.
  - (Table headings are placed above.)
- Graphs are numbered sequentially.
- Allowed to copy figures from the output of software.

- Grey scales + differentiable shapes
- Colors + differentiable shapes
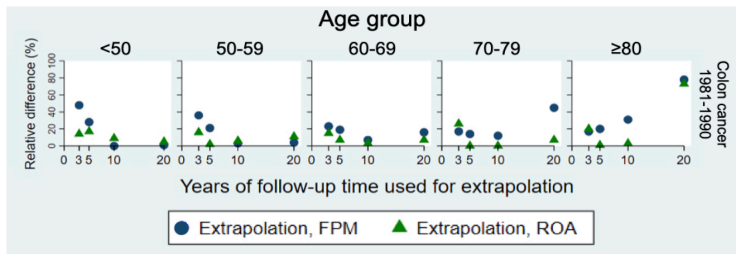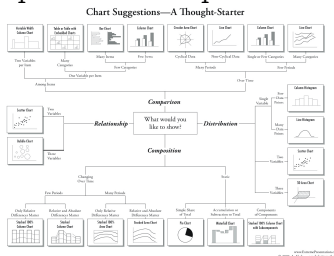- x-axis and y-axis are clearly defined with units.

Figure 6: Relative difference of different limited follow-up time used for extrapolation by age
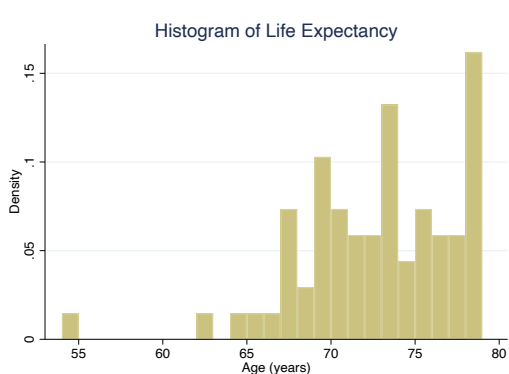
Graphs: Study map

Check the webpage: https://extremepresentation.com/tools/

Chart Suggestions—A Thought-Starter

Graphs: Histogram

Histogram depicts the distribution of data, where x-axis is usually a continuous variable.
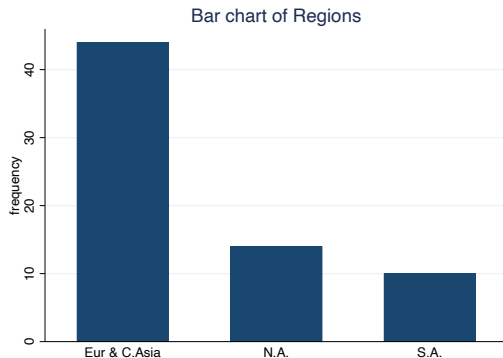
```
hist lexp, title("Histogram of Life Expectancy") ///
           xtitle(Age (years)) width(1) /// By each age
           graphregion(color(white)) //
```

Graphs: Bar chart

Bar chart shows the distribution of discrete (categorical) data.
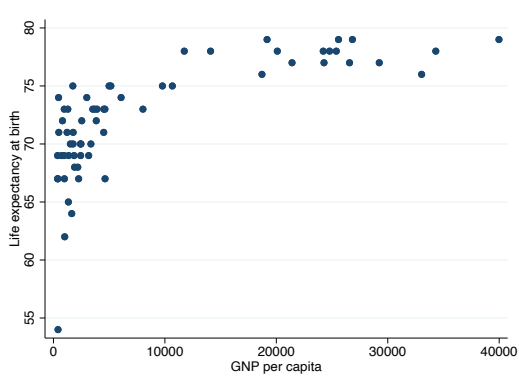
```
graph bar (count), over(region) ///
                title("Bar chart of Regions") ///
                graphregion(color(white)) //
```

Graphs: Scatter plot

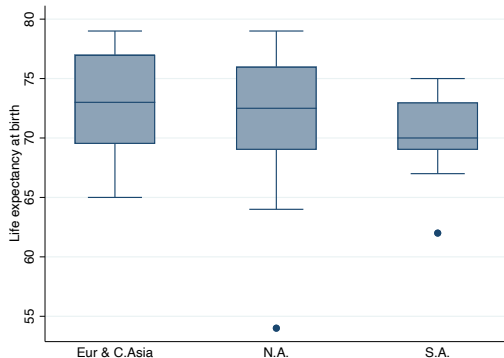Scatter plot demonstrates the relationship between two continuous variables.

```
twoway scatter lexp gnppc, graphregion(color(white))
```

Graphs: Box plot

Box plot summarises the distribution of the data, with the 25th, 50th, and 75th percentile and 1.5 IQR.
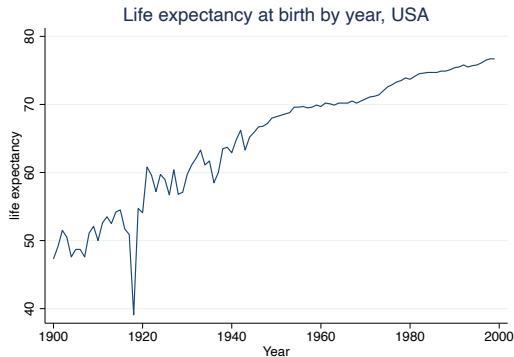
```
graph box lexp, over (region) ///
                graphregion(color(white))
```

Graphs: Line graph

Line graph functions similarly as scatter plots, with time as x-axis usually.

```
sysuse uslifeexp, clear
twoway line le year, title("Life expectancy at birth by year, USA")
       graphregion(color(white)) }
```
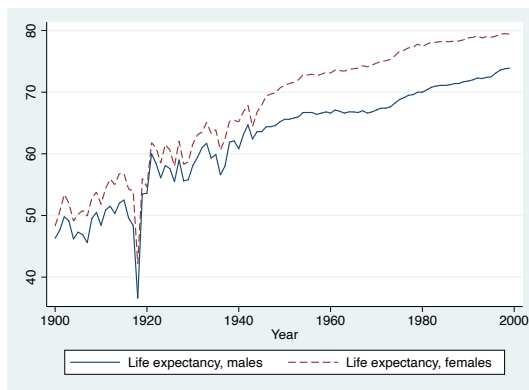
Graphs: Stratification

Data is already in separate columns. Or using by().
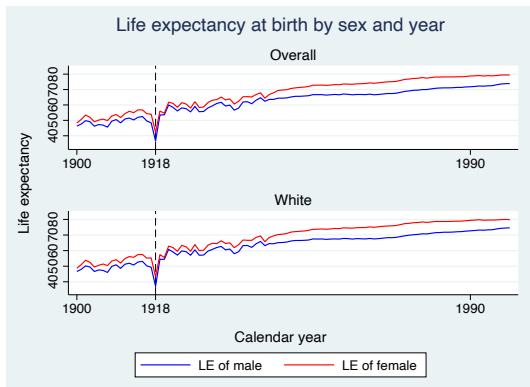Hint: by() is often used in individual-level data.
Btw, why did I use one solid line and the other dash line here?

Graphs: Putting graphs together

`grc1leg2` plays the role in plotting graphs together.
Hint: `grc1leg2` is not a default Stata command. See `help grc1leg2` to install it.
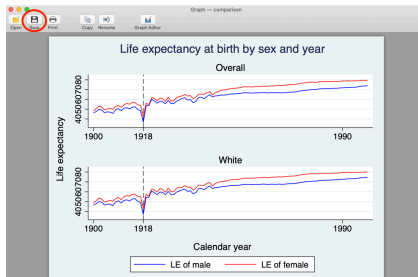
Graphs: Export

- A standard way:

```
graph export "location" /// assign the location
, as(pdf) name("")
```
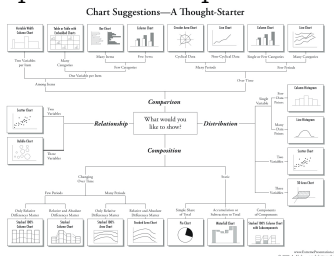
- An intuitive way:

- And then copy and paste the code back to the do-file.

Check the webpage: https://extremepresentation.com/tools/

References

1. Gordis L. *Epidemiology*. Philadelphia, PA: Elsevier/Saunders, 2014. ISBN 9781455737338.

2. David Goldsman PG. *A First Course in Probability and Statistics*. Georgia Institute of Technology, 2020.

3. Marcello Pagano KG. *Principles of Biostatistics*. Cengage Learning, Inc, 2000. ISBN 0534229026.

4. Erin Gabriel PF. Epidemiology PhD program, Karolinska Institutet, 2020.