Applied Epidemiology I: Graphs

Enoch Yi-Tung Chen

Department of Medical Epidemiology and Biostatistics, Karolinska Insitutet

February 8, 2021

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

Recapture: Risk difference

	Female (Exposed)	Male (Unexposed)	Total
shiba (Case)	2	2	4
guinea pig (Noncase)	2	1	3
Total	4	3	7

Epidemiologists love two by two tables!

- Risk difference between females having shiba and males having shiba $= \widehat{p_F} - \widehat{p_M} =$ $2/4 - 2/3- = -0.16667$
- Interpretation: Females have ~~16.67 % lower risk of having shiba~~ 16.67 less per 100 subjects of having shiba than males.

Recapture: Table or Graph?

Which one to show? [1]

Graph	**Table**
Visual	Written
Easier to understand	Requires more attention (decimals!)
General trends or comparisons	Specific or precise information
Easier to show time trend	Easier to the exact value

[1] Suggested reading: `https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/DataPresentation/DataPresentation2.html`

Basics: Bad examples

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

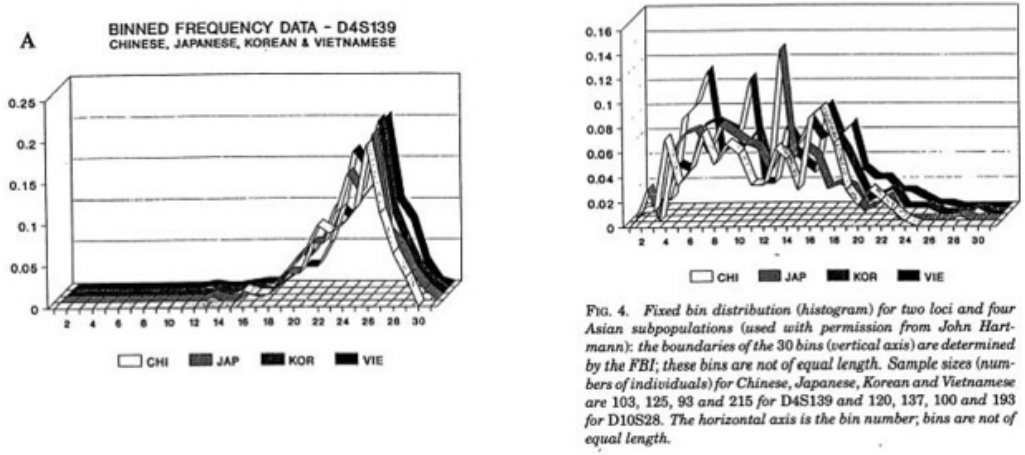
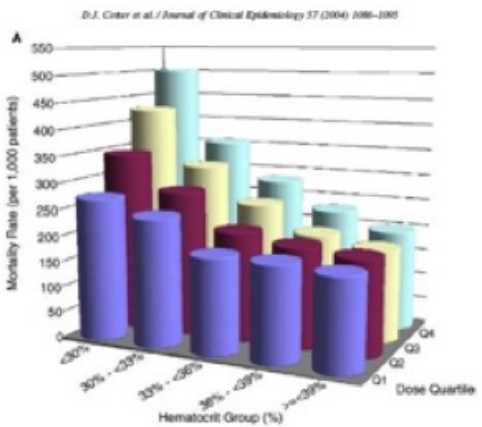
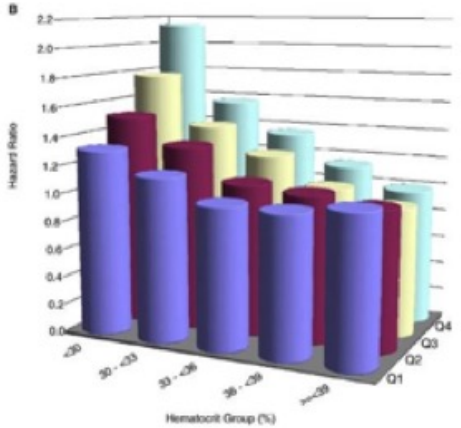


FIG. 4. *Fixed bin distribution (histogram) for two loci and four Asian subpopulations (used with permission from John Hartmann): the boundaries of the 30 bins (vertical axis) are determined by the FBI; these bins are not of equal length. Sample sizes (numbers of individuals) for Chinese, Japanese, Korean 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.*

Basics: Bad examples

Too fancy?





Basics: Bad examples

Insufficient info?

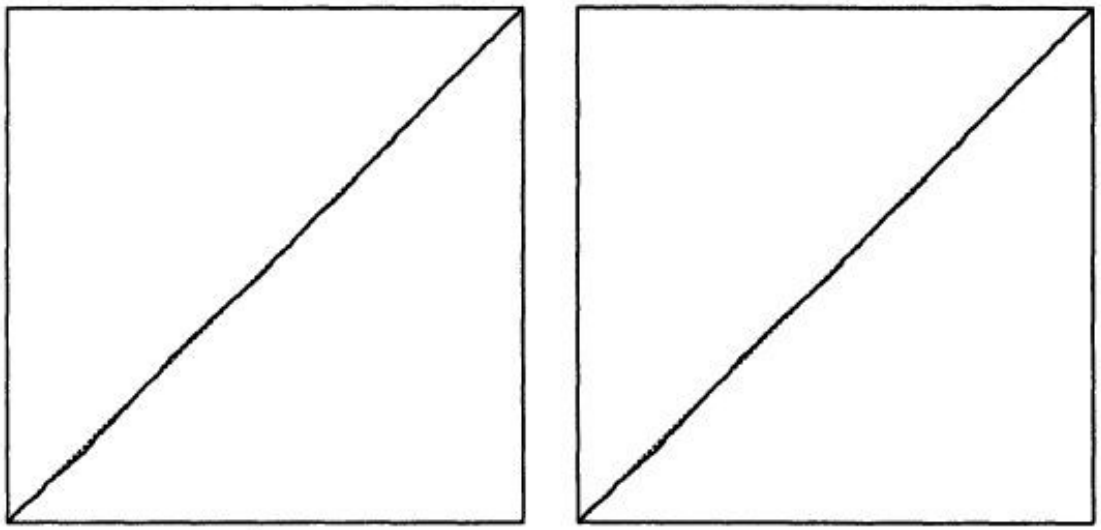
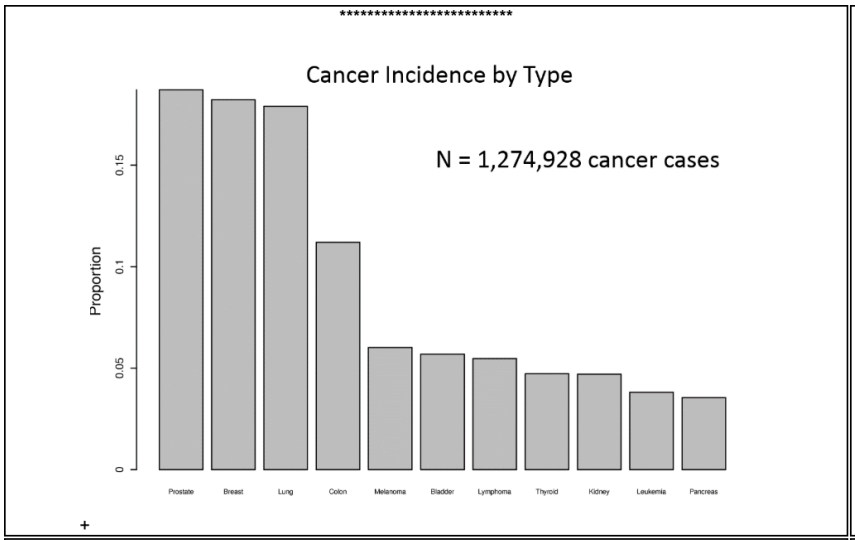
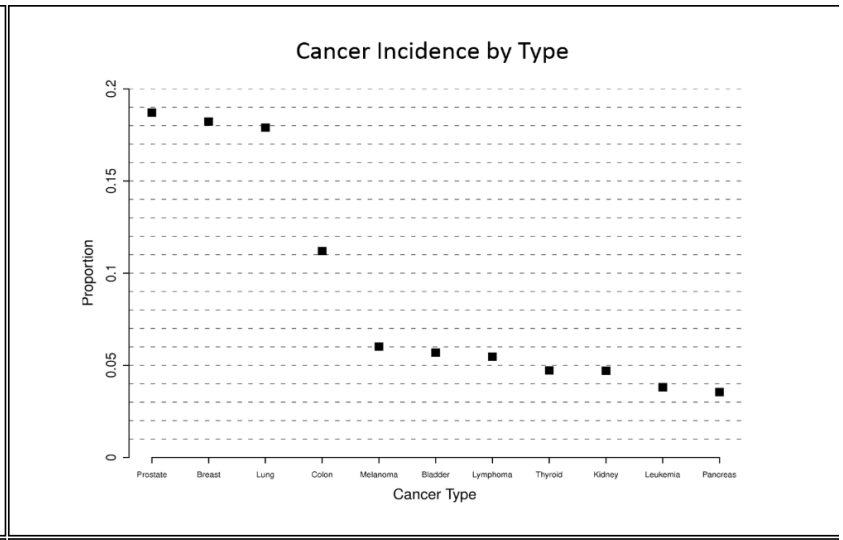


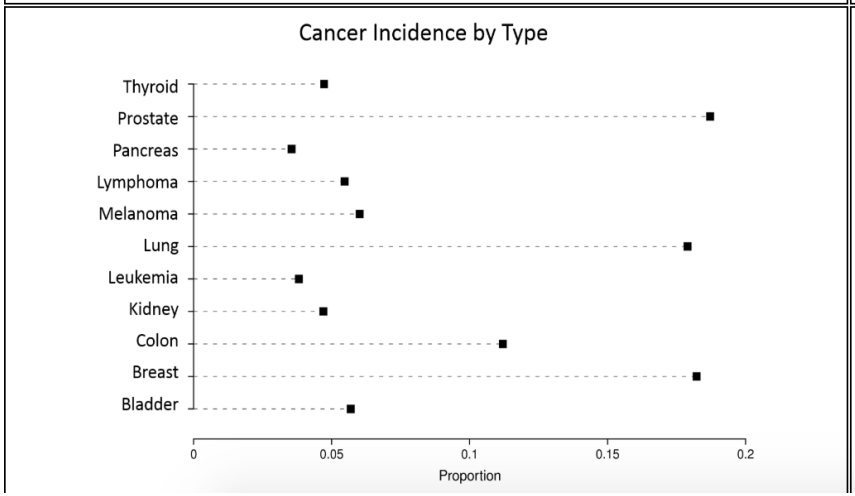
Figure 1. SRQ Plots of T_i/T_n (Vertical Axes) Against i/n (Horizontal Axes) for the Gibbs Sampler (a) and an Alternating Gibbs/Independence Sampler (b) for the Pump Failure Data Based on Runs of Length 5,000. Lines through the origin with unit slope are shown dashed; axis ranges are from 0 to 1 for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

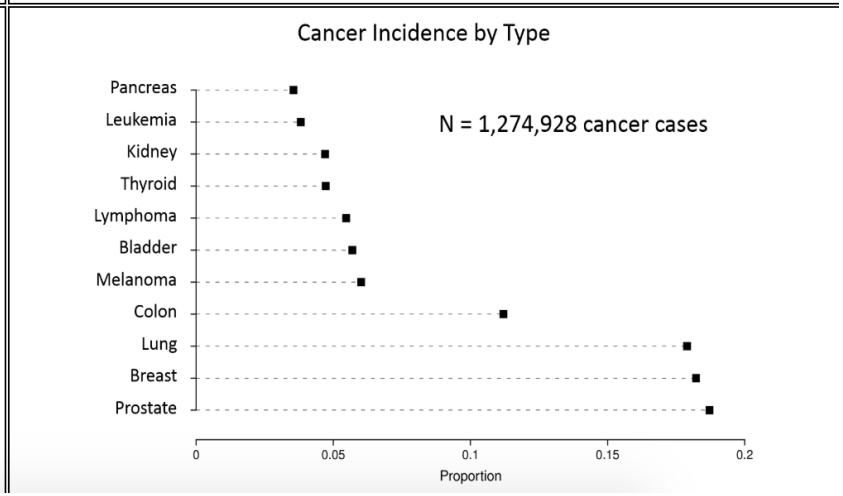
Basics: Bad examples

Sometimes there is no right nor wrong, it just depends on your interest.





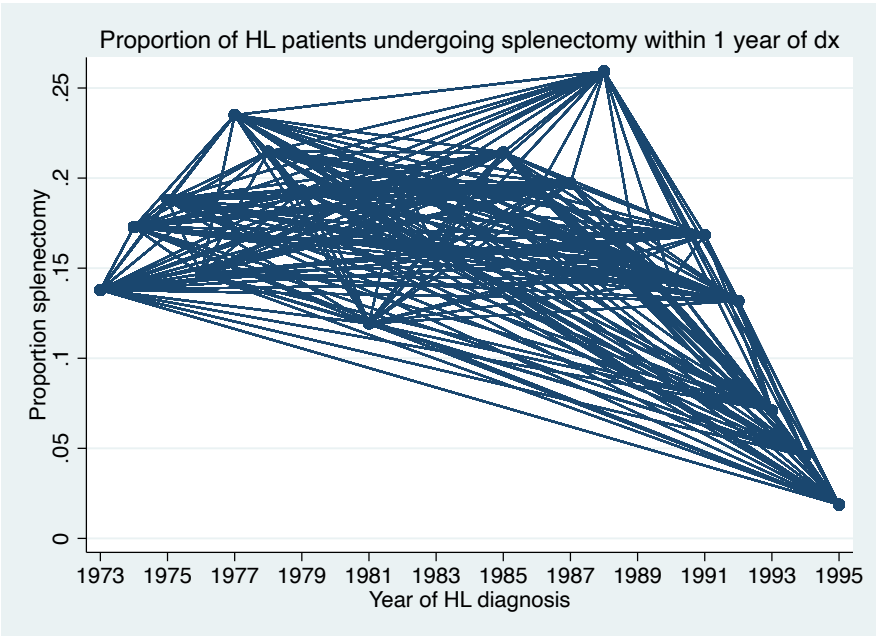




Basics: Learning from errors

Which part went wrong here?
Hint: something was missed in the code.

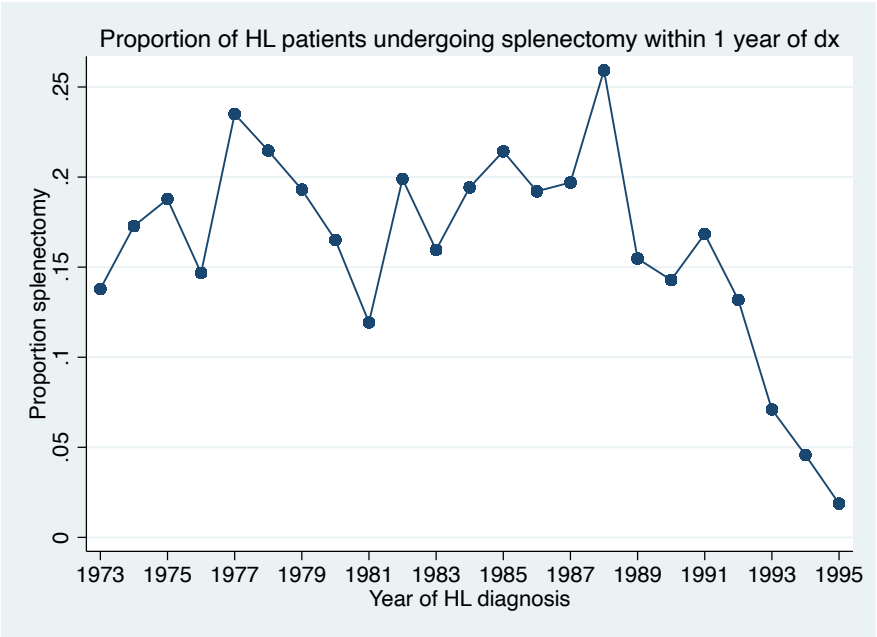
```
twoway connected prop diagyear, ///
subtitle("Proportion of HL patients") ///
ytitle(Proportion splenectomy) ///
xlabel(1973(2)1995)
```



Basics: 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
```



Basics: Basics of making graphs

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

Basics: Basics of making graphs

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

Basics: 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.

Basics: 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

Basics: 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.

Basics: 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.

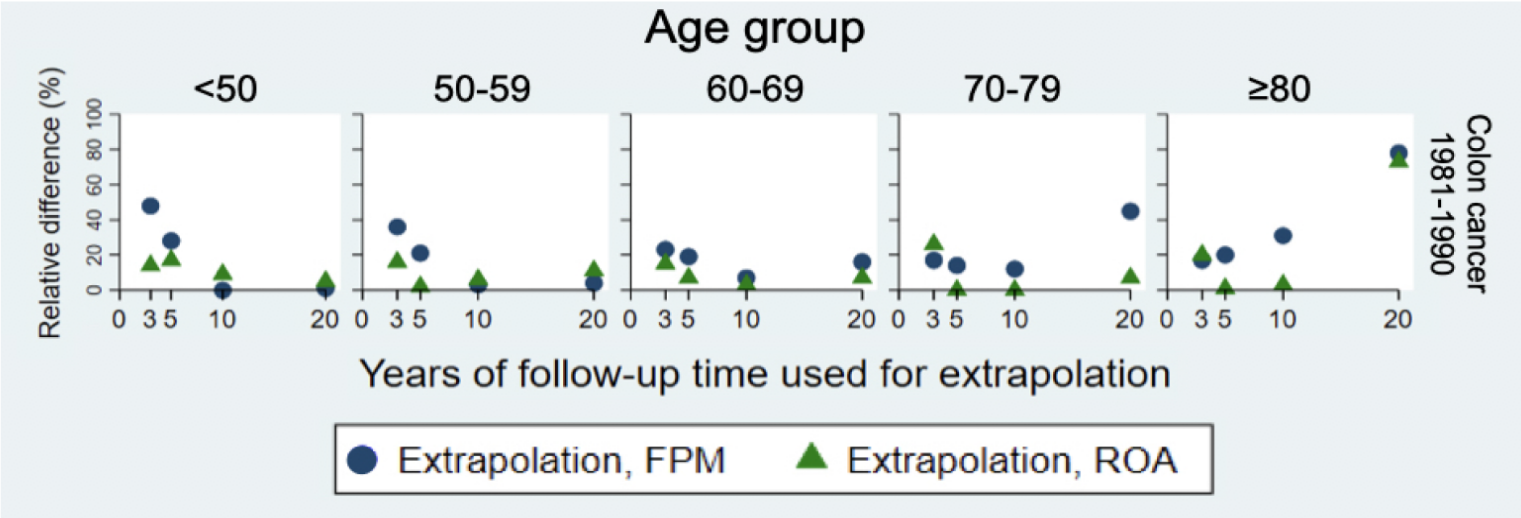
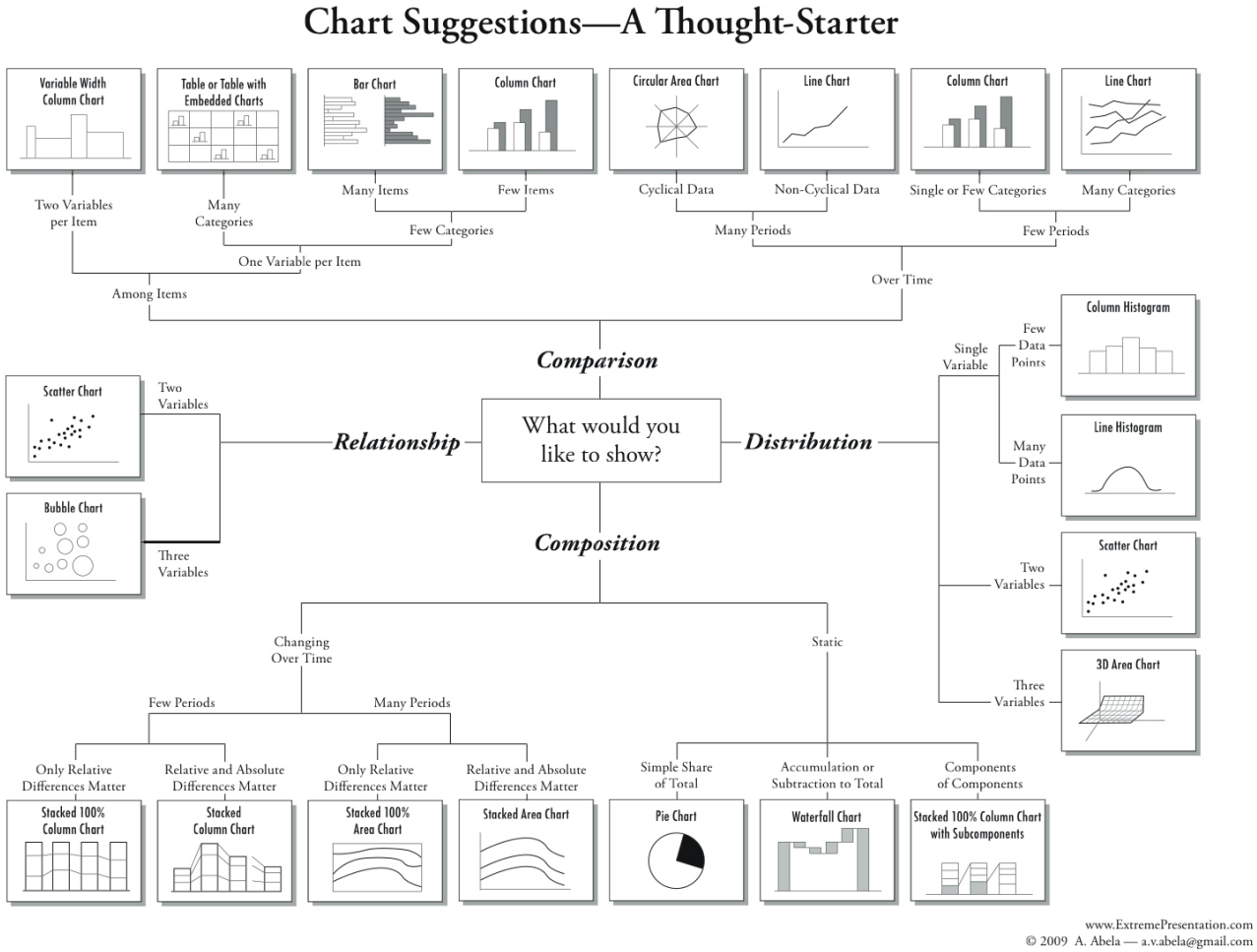


Figure 6: Relative difference of different limited follow-up time used for extrapolation by age group, colon cancer, diagnosis in 1981-1990, Sweden

Basics: Study map

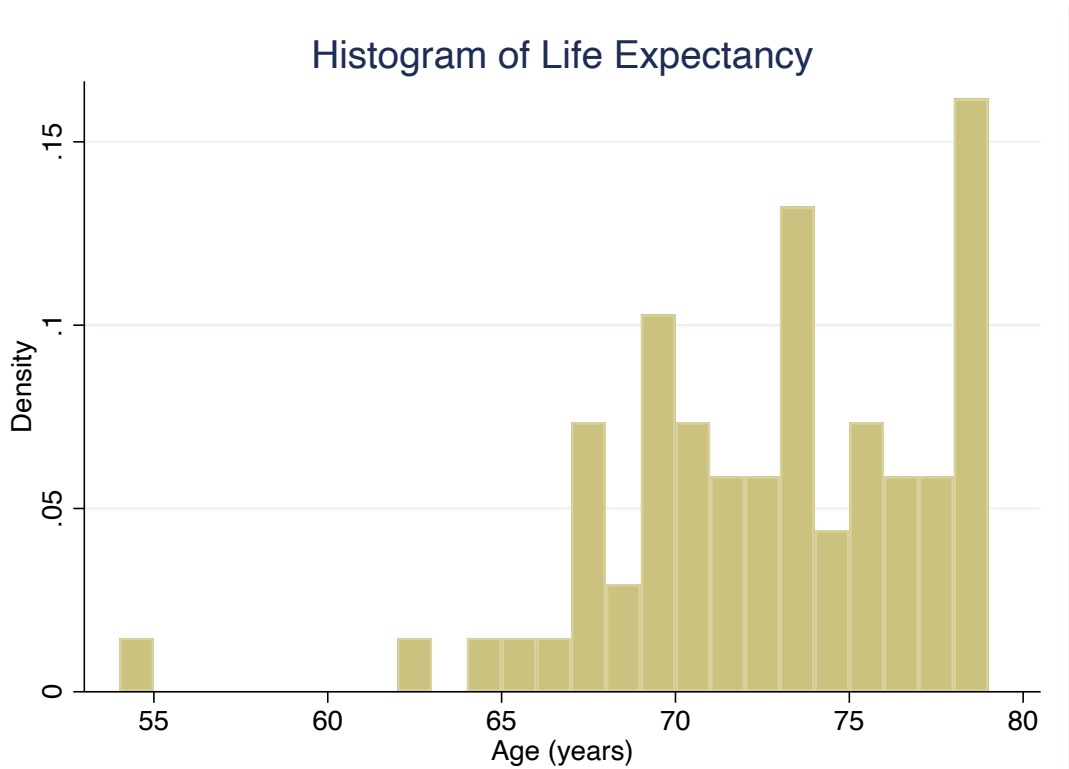
Check the webpage: `https://extremepresentation.com/tools/`



Types of graph: Histogram

Histogram depicts the distribution of data, where x-axis is usually a continuous variable.

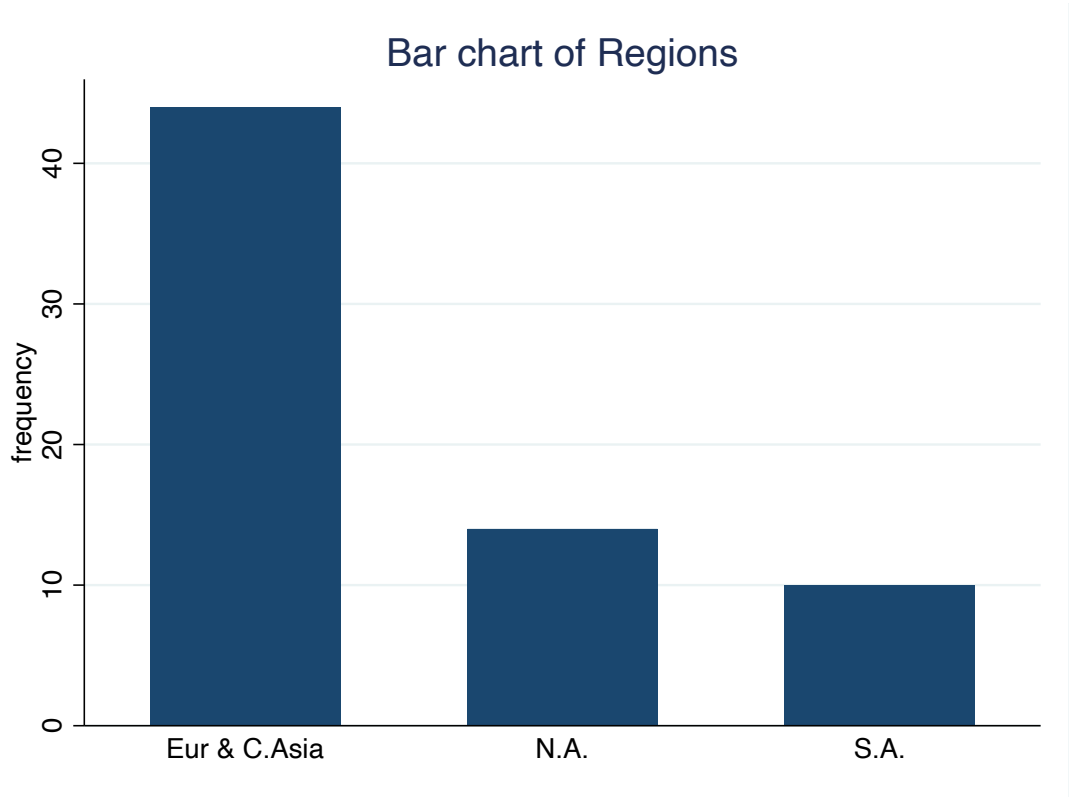
```
hist lexp, title("Histogram of Life Expectancy") ///
           xtitle(Age (years)) width(1) /// By each age
           graphregion(color(white)) //
```



Types of graph: Bar chart

Bar chart shows the distribution of discrete (categorical) data.

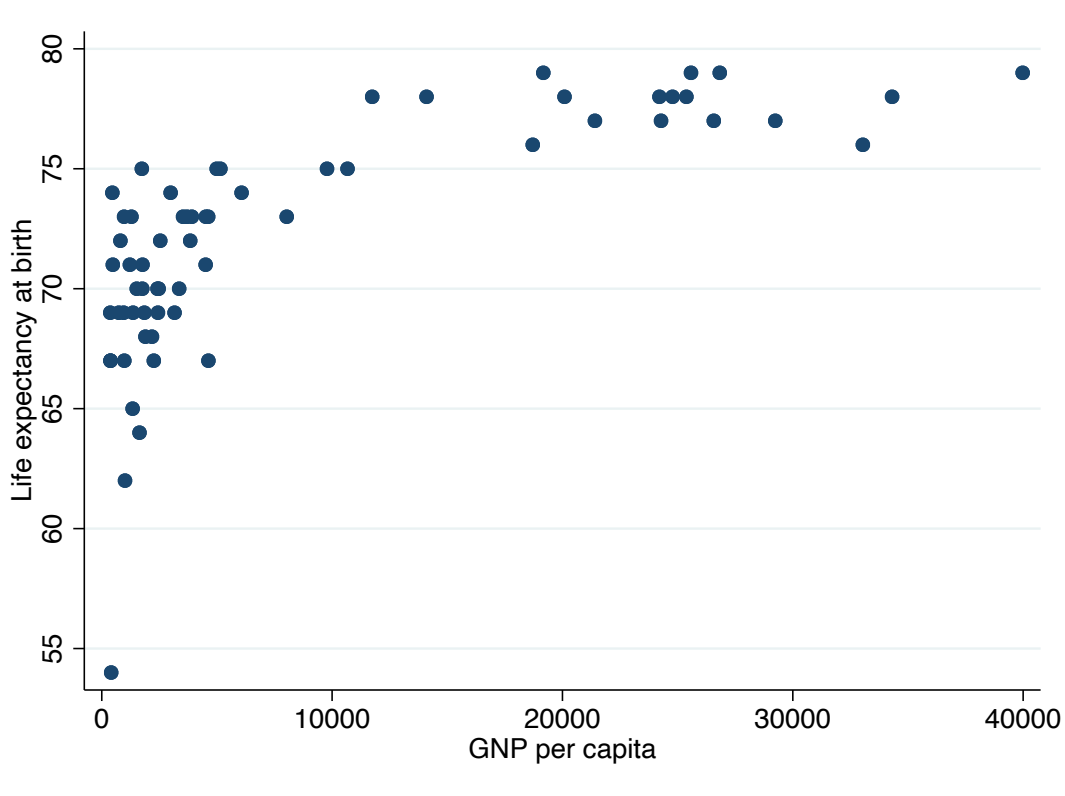
```
graph bar (count), over(region) ///
                   title("Bar chart of Regions") ///
                   graphregion(color(white)) //
```



Types of graph: Scatter plot

Scatter plot demonstrates the relationship between two continuous variables.

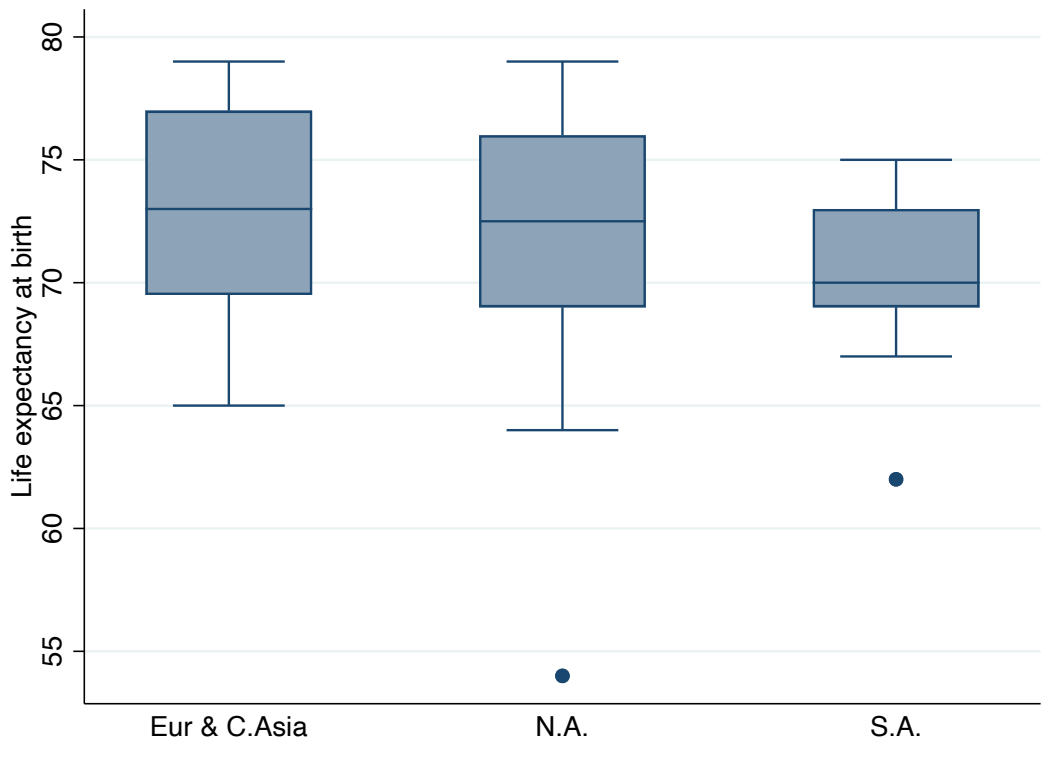
```
twoway scatter lexp gnppc, graphregion(color(white))
```



Types of graph: Box plot

Box plot summarises the distribution of the data, with the 25th, 50th, and 75th percentile and 1.5 IQR.

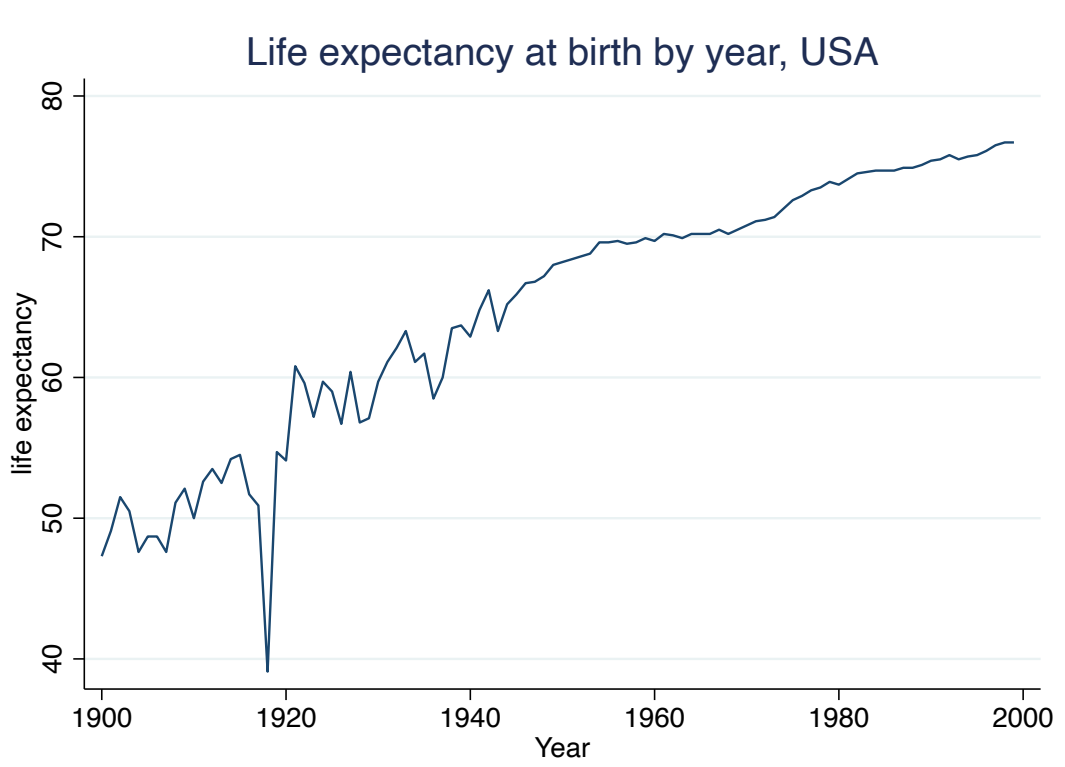
```
graph box lexp, over (region) ///
                graphregion(color(white))
```



Types of graph: Line graph

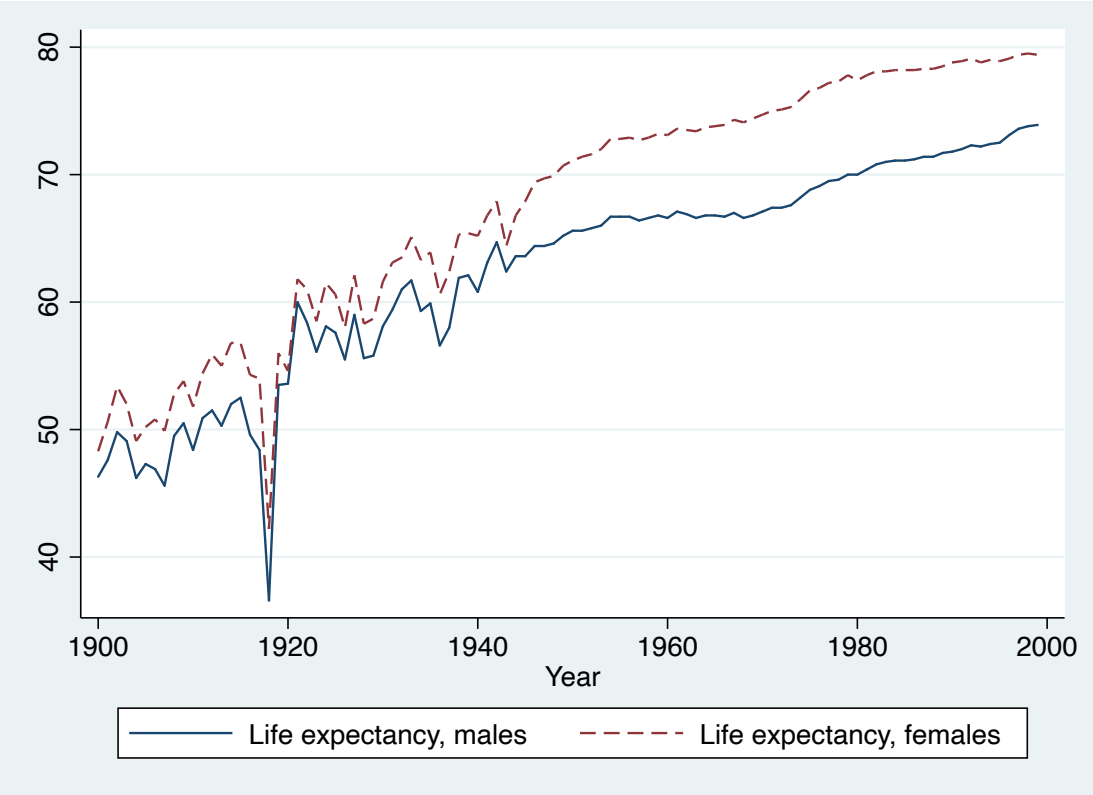
Line graph functions similarly as scatter plots, with time as x-axis usually.

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

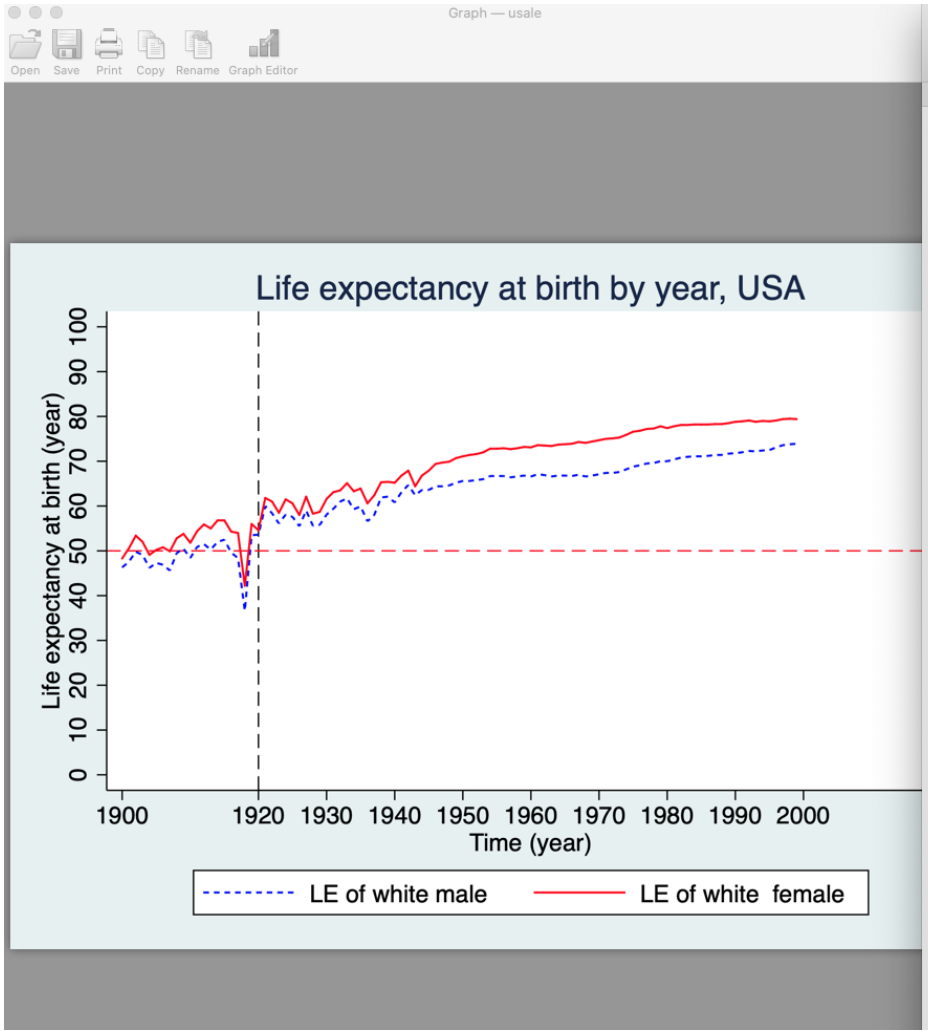


Types of graph: Stratification

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?



Powerful twoway: My most frequently used



graphs.do

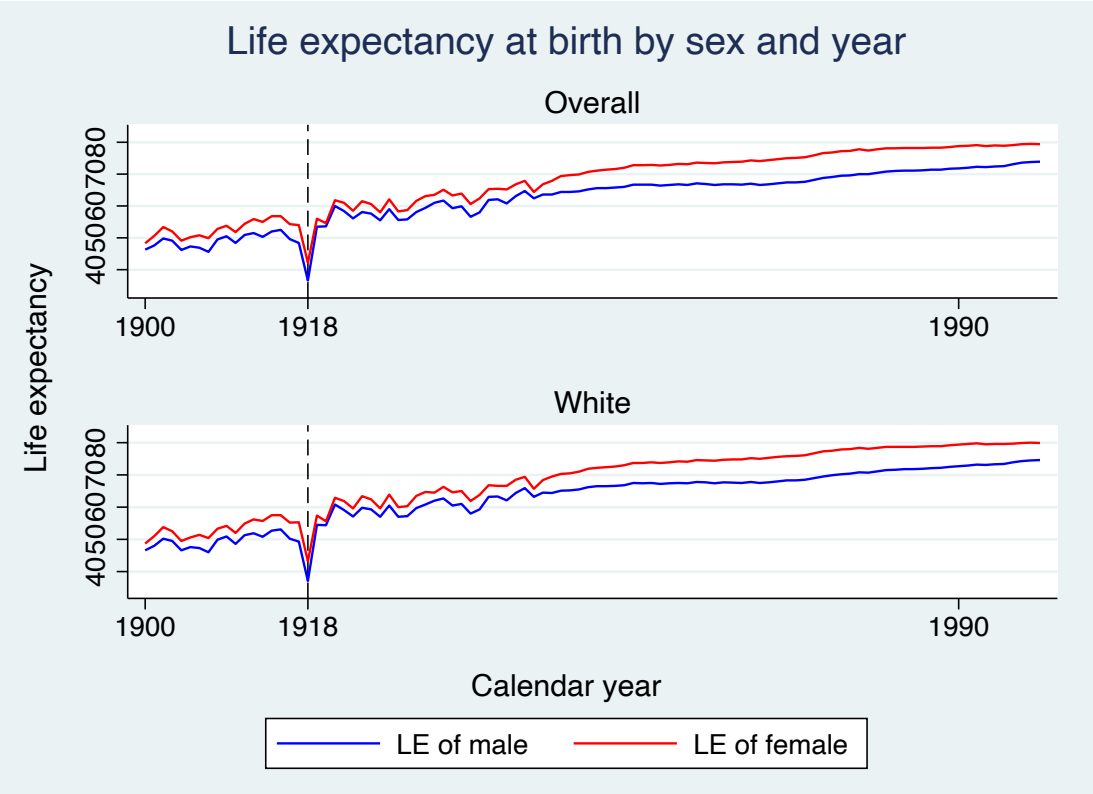
Open Save Print Find Show Zoom 166%

graphs.do

```
60
61  // Most frequently used: twoway
62  // Type -help twoway- to see the functions of twoway
63  twoway (line /*plotype*/ le_male/*y-axis*/ year/*x-axis*/, ///
64              lcolor(blue)/*line color*/ lpattern(shortdash) /*line pattern*/
65         (line le_female year, lcolor(red) lpattern(solid)), ///
66              ///
67         legend (order(1 "LE of male" 2 "LE of female")) ///
68              /// legend order by the same order above.
69              /// Make sure you get all the space, quotation marks right!
70              ///
71         xline(1920, lpattern(dash) lwidth(thin) lcolor(black))        ///
72              /// Add vertical line x=1920
73         xscale(range(1900(10)2020)) ///
74              /// xscale's main function is to change the scale of the axis
75         xlabel(1900 1920(10)2000) ///
76              /// xlabel would label the appointed values
77         xtitle("Time (year)") ///
78              /// Put name + unit
79              ///
80         yline(50,lpattern(dash) lwidth(thin) lcolor(red)) ///
81              /// Add horizontal line y = 50
82         yscale(range(0(10)100)) ///
83              /// Set the scale from 0 to 100
84         ylabel(0(10)100, nogrid) ///
85              /// You don't need to put range in label()
86              /// Take away grid line by adding -nogrid-
87         ytitle("Life expectancy at birth (year)") ////
88              /// Put name + unit
89              ///
90         name("usale", replace) ///
91              /// Give the name for furtheruse
92         title("Life expectancy at birth by year, USA") ///
93              /// Main title
94         saving( ,replace)
95              /// Save as gph, but for other file types: -graph export -
96
```

Powerful twoway: 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.

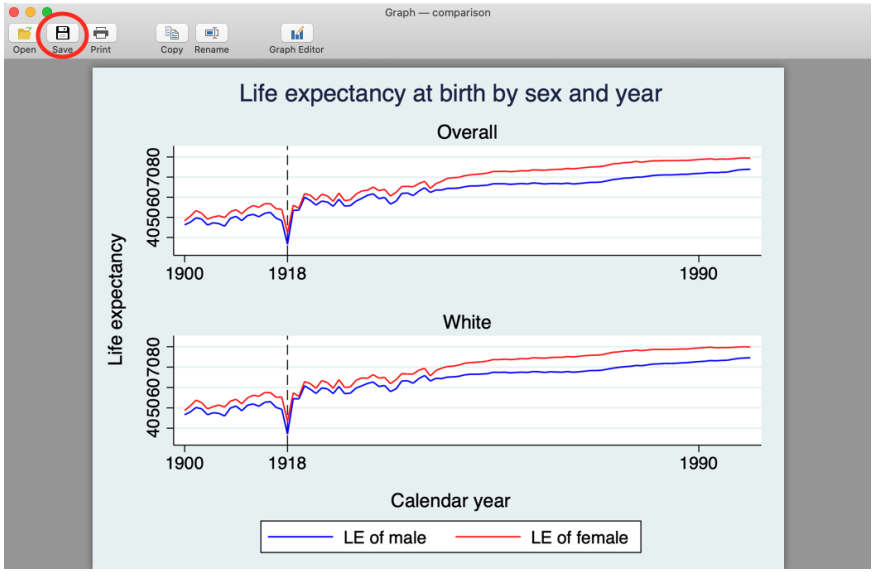


Powerful twoway: Export

- A standard way:

```
graph export "location" /// assign the location
, as(pdf) name("")
```

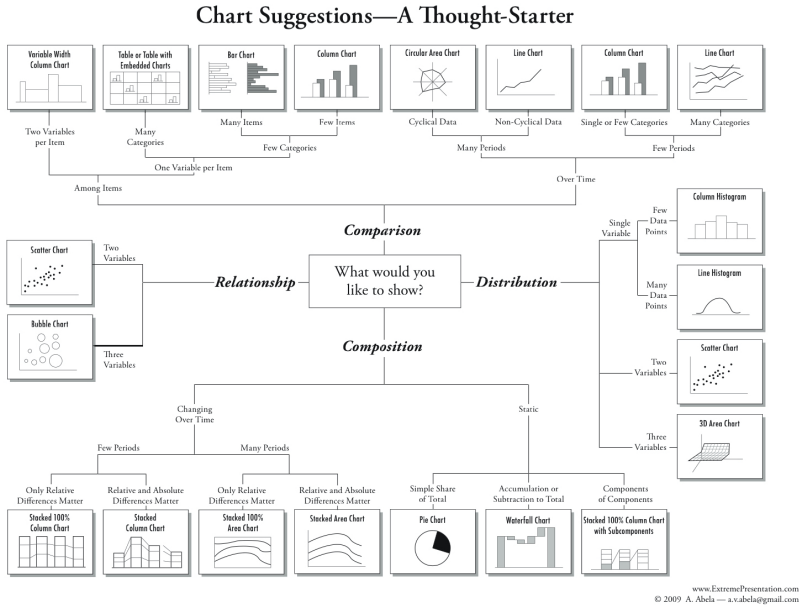
- An intuitive way:



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

Study map

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



Other resources:

1. UCLA. GRAPHICS:OVERVIEW OF TWOWAY PLOTS — STATA LEARNING MODULES. `https://stats.idre.ucla.edu/stata/modules/graph8/twoway/`
2. Infogram. How to Choose the Right Chart for Your Data. `https://infogram.com/page/choose-the-right-chart-data-visualization`
3. Looker. How to choose the best chart or graph for your data. `https://looker.com/blog/different-types-graphs-charts-uses`

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.