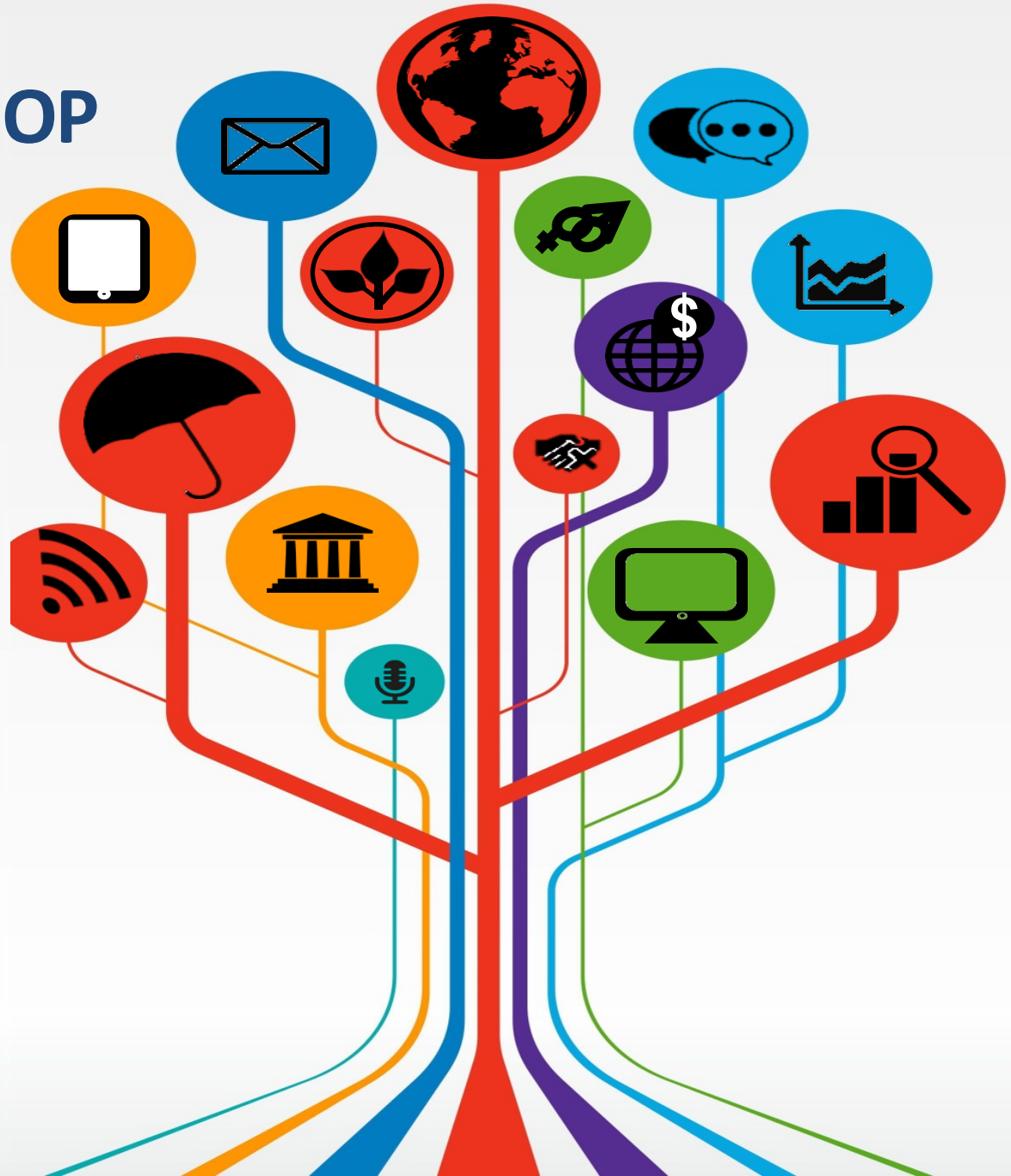
FIELD COORDINATOR WORKSHOP

Manage Successful Impact Evaluations

18 - 22 JUNE 2018
WASHINGTON, DC



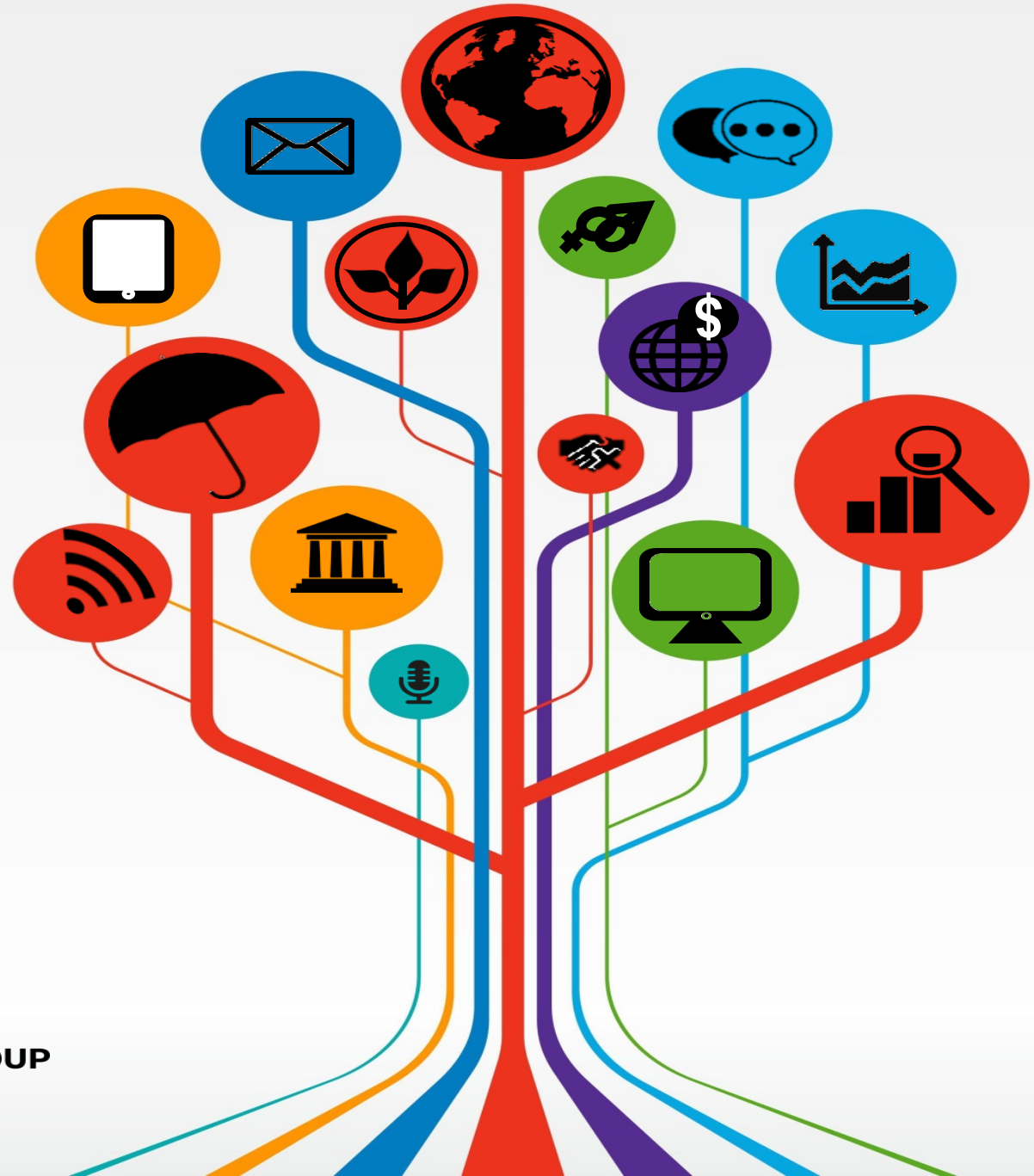
i2i DIME
TRANSFORM DEVELOPMENT

WORLD BANK GROUP

Data Visualization

Mrijan Rimal & Kristoffer Bjärkefur

June 21, 2018



i2i DIME
TRANSFORM DEVELOPMENT

WORLD BANK GROUP

Tables give all the details

- What's happening in this regression table? What's important?

TABLE 3—MEASURES OF ACCESS TO CARE JUST BEFORE 65 AND ESTIMATED DISCONTINUITIES AT 65

	1997–2003 NHIS				1992–2003 NHIS			
	Delayed care last year		Did not get care last year		Saw doctor last year		Hospital stay last year	
	Age 63–64 (1)	RD at 65 (2)	Age 63–64 (3)	RD at 65 (4)	Age 63–64 (5)	RD at 65 (6)	Age 63–64 (7)	RD at 65 (8)
Overall sample	7.2	−1.8 (0.4)	4.9	−1.3 (0.3)	84.8	1.3 (0.7)	11.8	1.2 (0.4)
Classified by ethnicity and education:								
White non-Hispanic:								
High school dropout	11.6	−1.5 (1.1)	7.9	−0.2 (1.0)	81.7	3.1 (1.3)	14.4	1.6 (1.3)
High school graduate	7.1	0.3 (2.8)	5.5	−1.3 (2.8)	85.1	−0.4 (1.5)	12.0	0.3 (0.7)
At least some college	6.0	−1.5 (0.4)	3.7	−1.4 (0.3)	87.6	0.0 (1.3)	9.8	2.1 (0.7)
Minority:								
High school dropout	13.6	−5.3 (1.0)	11.7	−4.2 (0.9)	80.2	5.0 (2.2)	14.5	0.0 (1.4)
High school graduate	4.3	−3.8 (3.2)	1.2	1.5 (3.7)	84.8	1.9 (2.7)	11.4	1.8 (1.4)
At least some college	5.4	−0.6 (1.1)	4.8	−0.2 (0.8)	85.0	3.7 (3.9)	9.5	0.7 (2.0)
Classified by ethnicity only:								
White non-Hispanic	6.9	−1.6 (0.4)	4.4	−1.2 (0.3)	85.3	0.6 (0.8)	11.6	1.3 (0.5)
Black non-Hispanic (all)	7.3	−1.9 (1.1)	6.4	−0.3 (1.1)	84.2	3.6 (1.9)	14.4	0.5 (1.1)
Hispanic (all)	11.1	−4.9 (0.8)	9.3	−3.8 (0.7)	79.4	8.2 (0.8)	11.8	1.0 (1.6)

Note: Entries in odd numbered columns are mean of variable in column heading among people ages 63–64. Entries in even numbered columns are estimated regression discontinuities at age 65, from models that include linear control for age interacted with dummy for age 65 or older (columns 2 and 4) or quadratic control for age, interacted with dummy for age 65 and older (columns 6 and 8). Other controls in models include indicators for gender, race/ethnicity, education, region, and sample year. Sample in columns 1–4 is pooled 1997–2003 NHIS. Sample in columns 5–8 is pooled 1992–2003 NHIS. Samples for regression models include people ages 55–75 only. Standard errors (in parentheses) are clustered by quarter of age.

But figures *tell the story*

- This is the data that generates those estimates.
- You can see exactly what is happening very quickly!

Even more importantly:

You don't have to look for it.
The eye is drawn to the story!

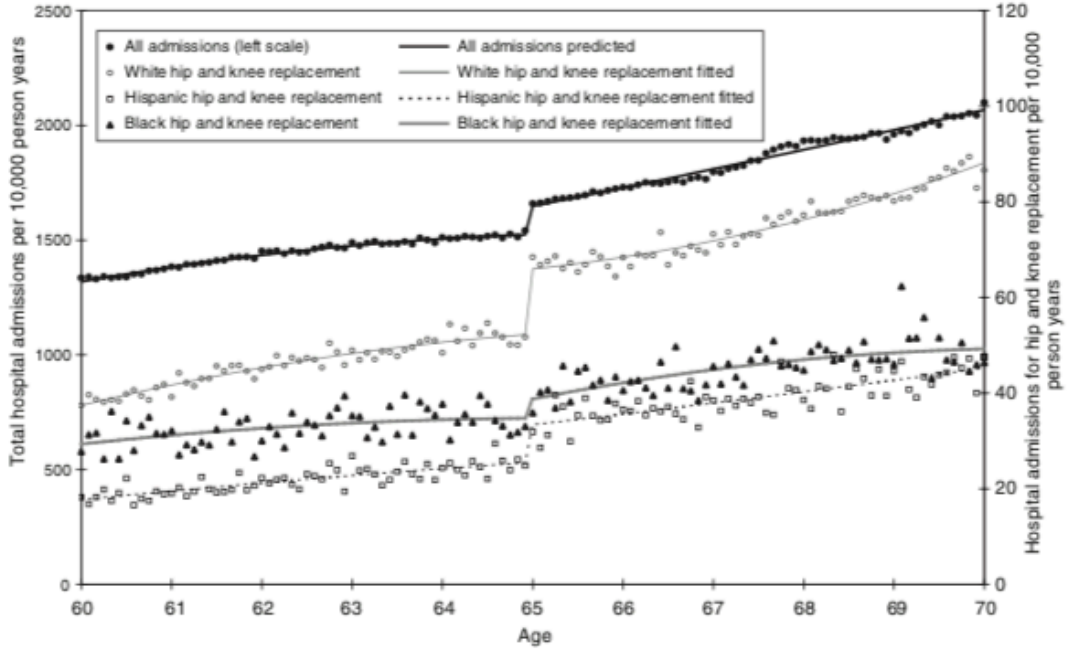
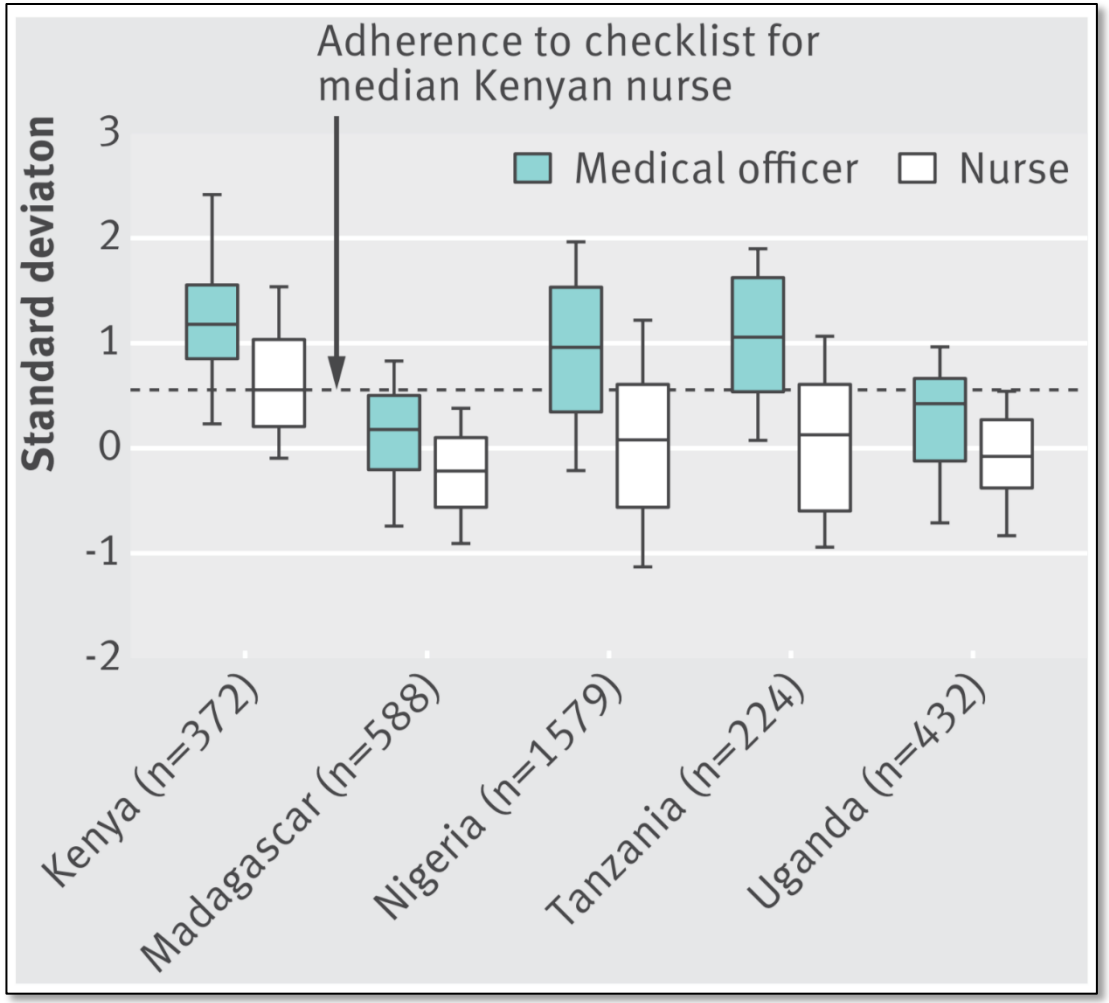


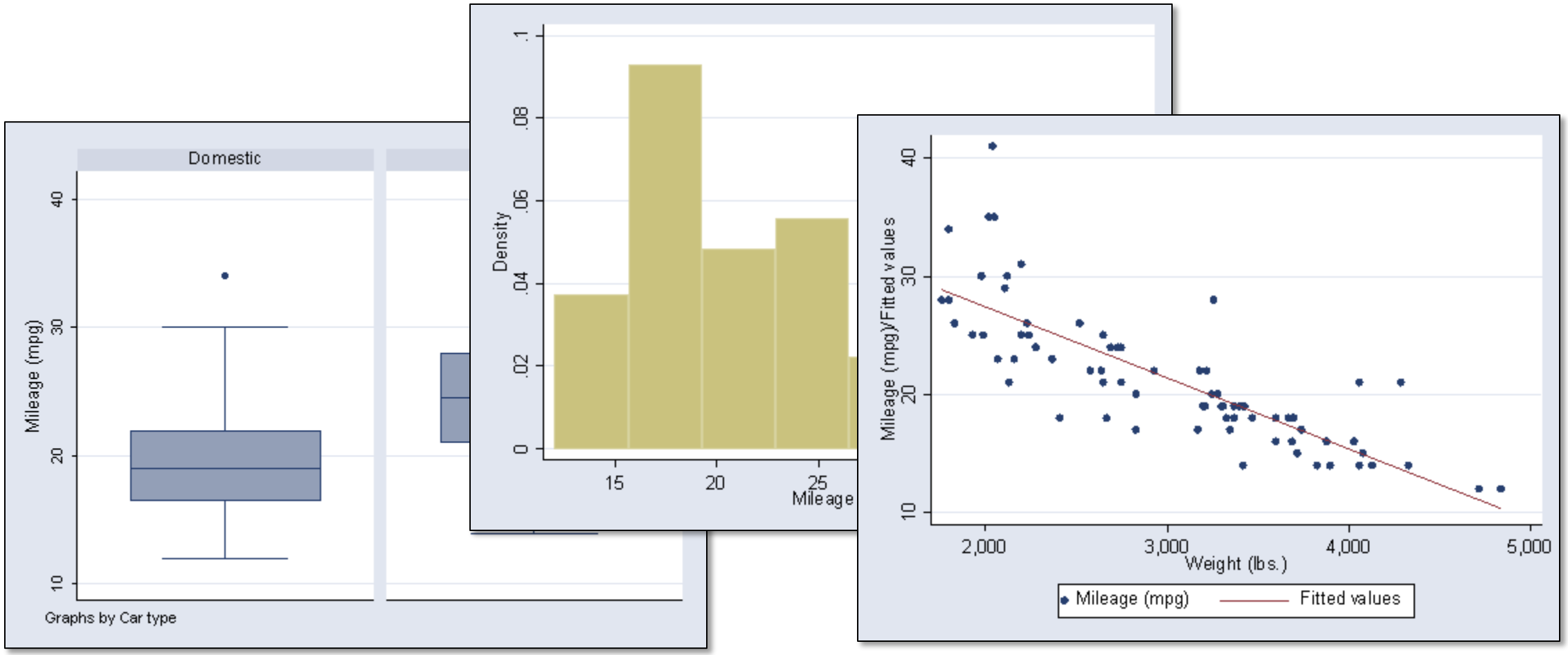
FIGURE 3. HOSPITAL ADMISSION RATES BY RACE/ETHNICITY

Examples: comparing means

- What is the main story in this graph?
- We need more context to say something detailed about this, but what has the person creating the graph highlighted for us?



Stata default graphs



Stata has three core built-in graph functions

[*graph graphtype*]

- graphs which plot one or more variables *on one axis*

[*twoway graphtype*]

- graphs which plot two variables *together on an x,y axis*

[*histogram*] , [*kdensity*] , [*lowess*]

- Essential *distributional* commands

The other graph commands are implemented in terms of **graph**, which provides the following capabilities:

Command	Description
graph bar	bar charts
graph pie	pie charts
graph dot	dot charts
graph matrix	scatterplot matrices
graph twoway	twoway (*y*-*x*) graphs, including
graph twoway scatter	scatterplots
graph twoway line	line plots
graph twoway function	function plots
graph twoway histogram	histograms
graph twoway *	more

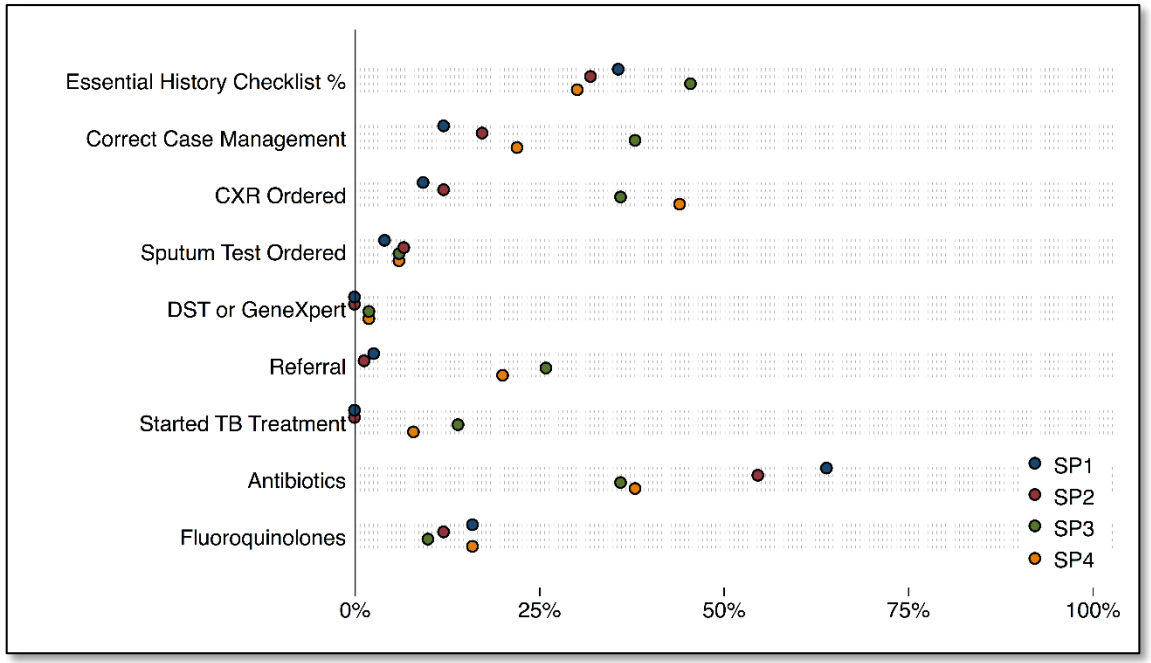
Smoothing and densities:

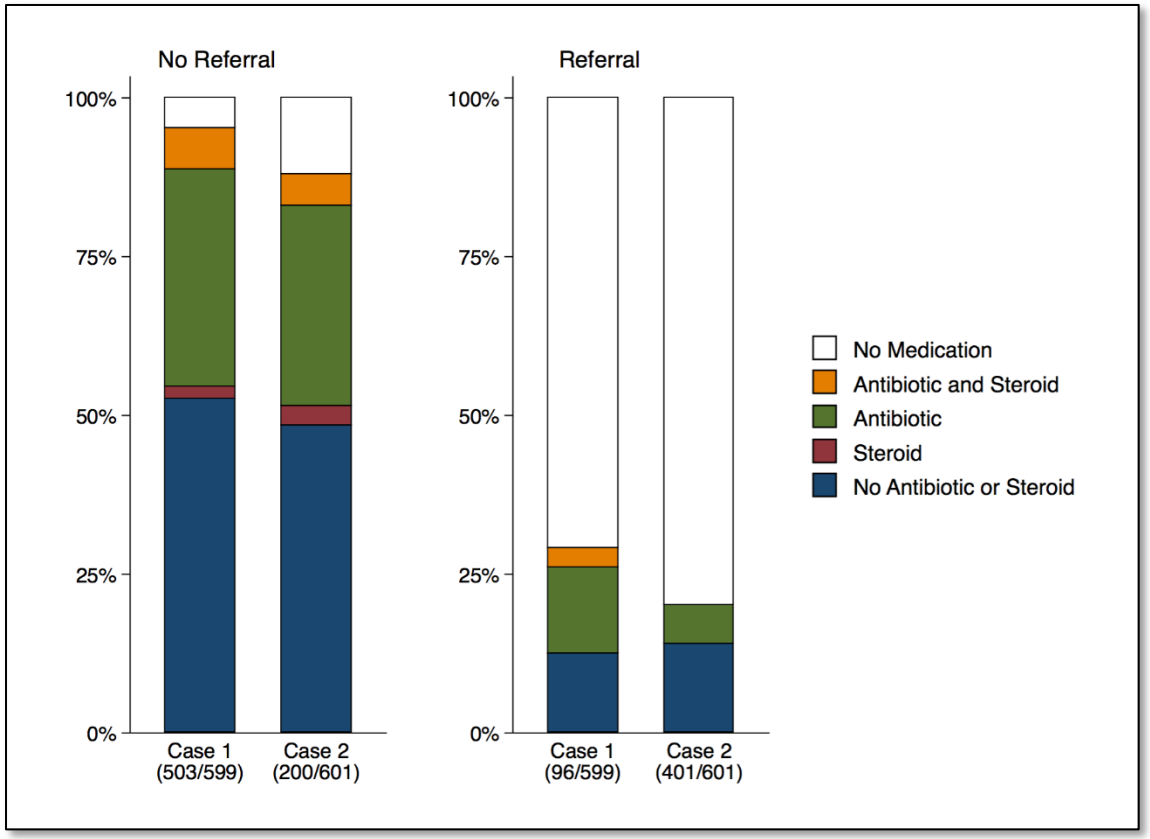
Command	Description
kdensity	kernel density estimation, univariate
lowess	lowess smoothing
lpoly	local polynomial smoothing

Task 1 - Histogram

1. Run the code for the histogram for variable *ag_16_x_16_1*
2. What can you tell from the graph?
3. Exclude outliers and run the graph again
   - You can exclude outliers by using *if ag_16_x_16_1 <* and add a value of your choice
4. After you have experimented run the histogram that we have prepared.
   - See how much you can change the look of the graph

Oneway [*graph*] plots can be very informative





https://github.com/qutubproject/lancetid2015/blob/master/tables_figures.do (Figure S1)

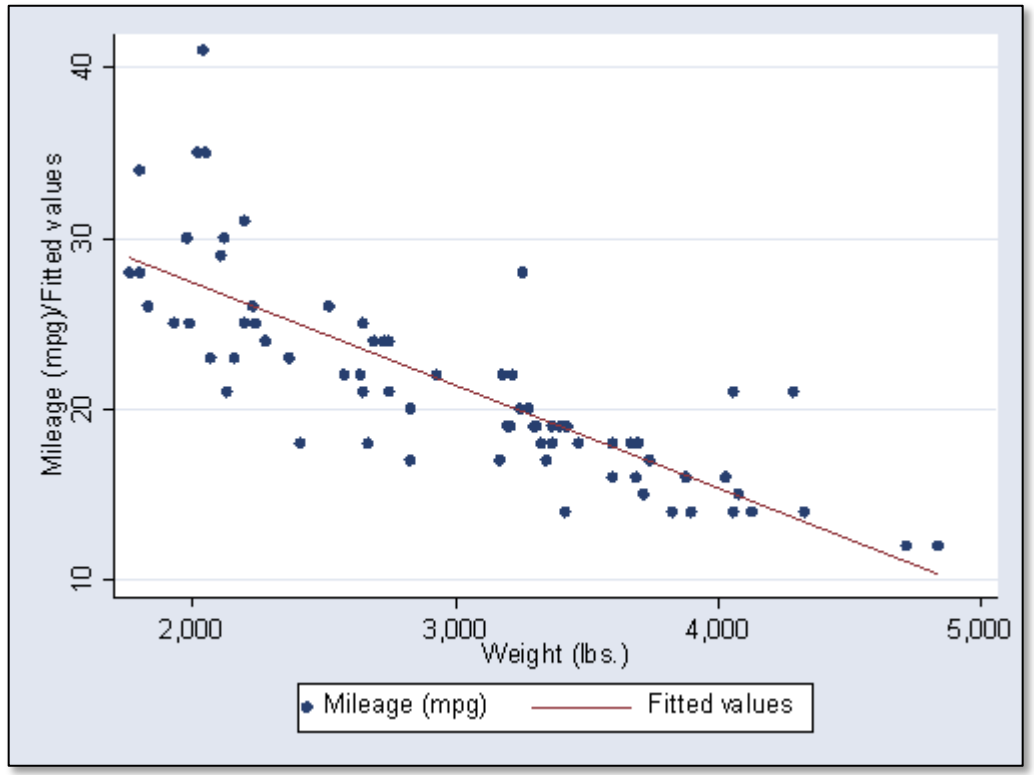
https://github.com/qutubproject/lancetid2016/blob/master/tables_figures.do (Figure 2)

Task 2 – Oneway Graph

1. Run the code *tab gr_16 ag_17_x_16_1 , row m*
   - This table shows differences in usage of hired labor across the LWH Cooperative groups, is it easy to get an overview?
2. Run the first bar graph in section Task 2.
   - This is the avarage of all observations. In oneway graphs we can usually add *by()* to show differences between groups.
3. Run the second bar graph where the option *by()* is used
   - To see the labels better we should make the bars horizontal.
4. Run the third bar graph where this and more is fixed

[*twoway*] graphs

- Each point in the graph represents a combination of the y-axis and the x-axis
- Can be many types of graphs. This is a scatter plot, but it can be areas, lines, bars, etc.



Multiple twoway graphs in one graph

- You can add multiple graphs in the same graph and format the data points slightly differently
- Syntax:

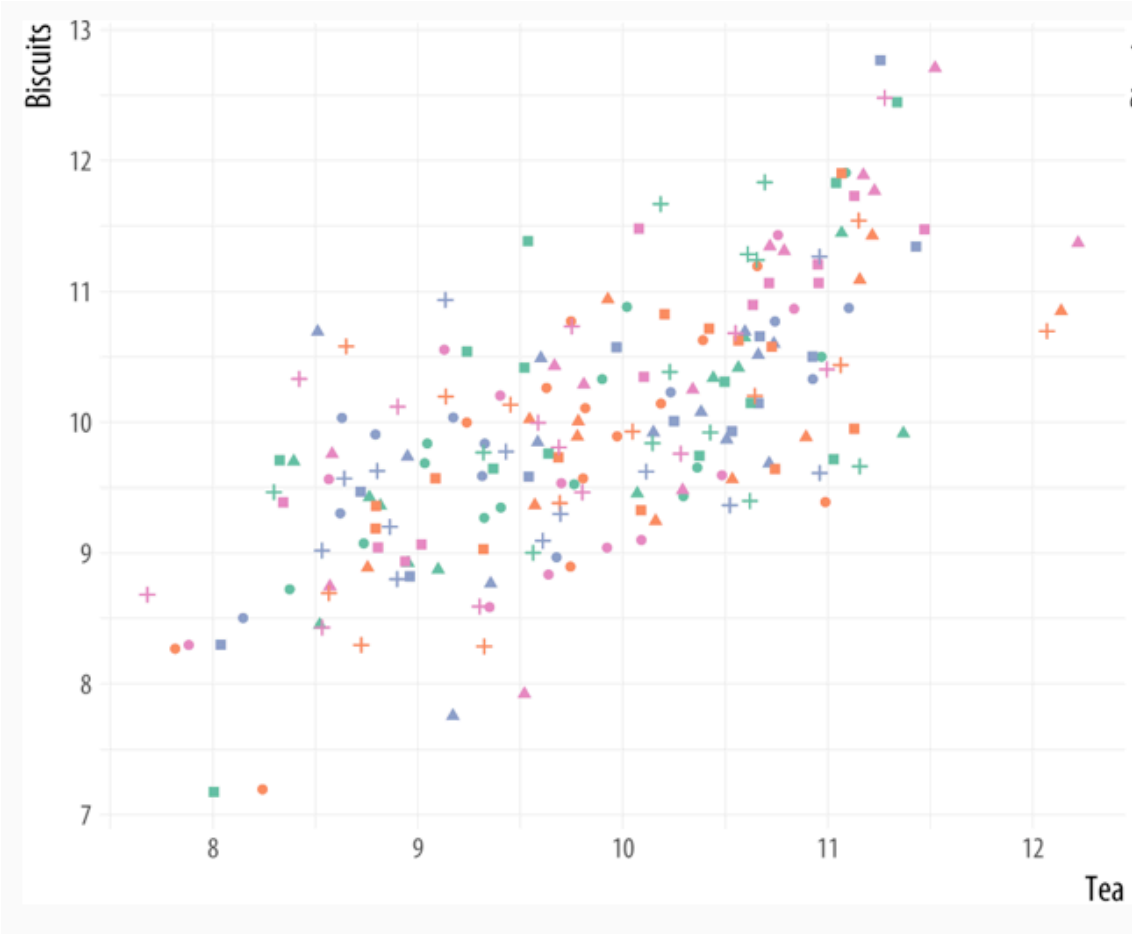
  [*tw ///*

  (type var1 var2 , opts) ///

  (type var3 var4 , opts) ///

  , opts]

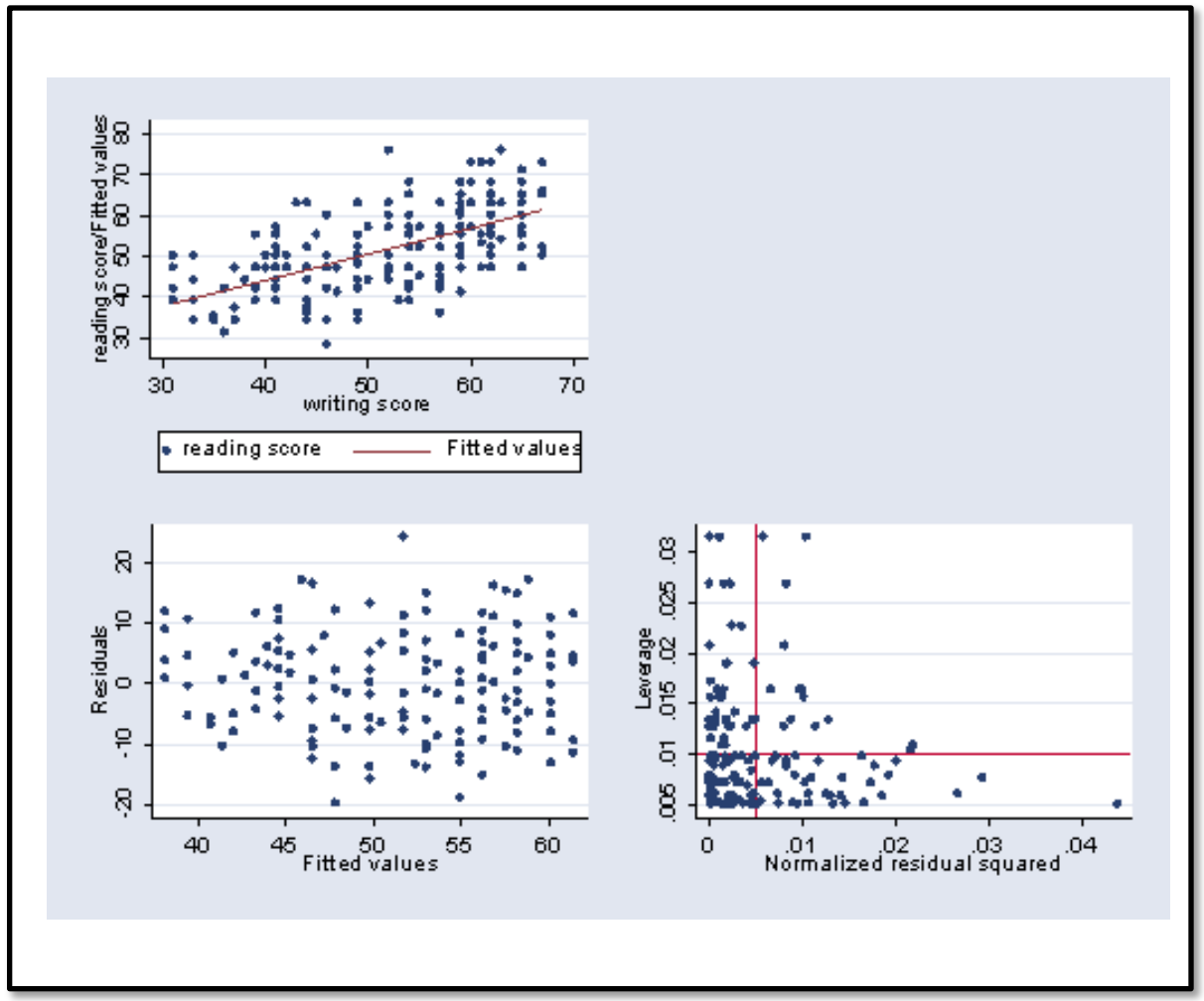
- More on this in Track 2



Task 3 – twoway graphs

1. Run twoway graph 1. This shows the relation between days spent on land preparation and amount spent on hired labor for control observations
2. In twoway graph 2 we add a line representing that weighted mean or spending on hired labor for each level of days spent on land preparation. The line is added as a second graph
3. In twoway graph 3, do the same graph but for treatment observations
4. Put both graphs for both control and treatment on the same graph in twoway graph 4. We have added more formatting options to this graph. This is four graphs in one graph

Graphs can be combined and exported



graph export ///

"filename" /// (.png or .eps)

, replace

With .png, specify "width(1000)" for higher resolution

.eps files can scale to any size on most modern software (but hard to preview on older systems)

Task 4 – Combining Graphs

The last graph for each task were saved to your folder. We can combine these graphs to one graph that we then can save and import to our report

1. Add the names of the graphs from task 2 and 3 and run the *graph combine* command.
2. Run the *graph export* command that is able to save the graph in many more formats than the .gph format

DIME Resources (please contribute!)

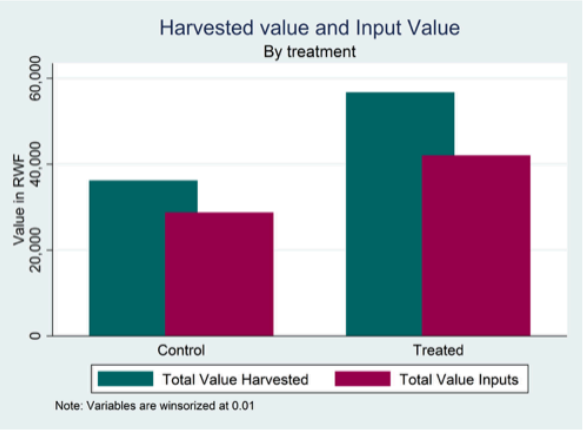
https://worldbank.github.io/Stata-IE-Visual-Library/

https://worldbank.github.io/stata/

IE Visual Library in Stata

Bar plots

Bar plot of two variables by treatment




```
# d;
graph bar w_total_val_harvested_a w_total_val_inputs_a,
          over(treated)
          bargap(-30)
          legend(label(1 "Total Value Harvested")
                         label(2 "Total Value Inputs"))
          bar (1, color("0 102 102") )
          bar (2, color("153  0 76") )
          ytitle ("Value in RWF")
          title  ("Harvested value and Input Value")
          subtitle ("By treatment")
          note ("Note: Variables are winsorized at 0.01");
# d cr
```

stata

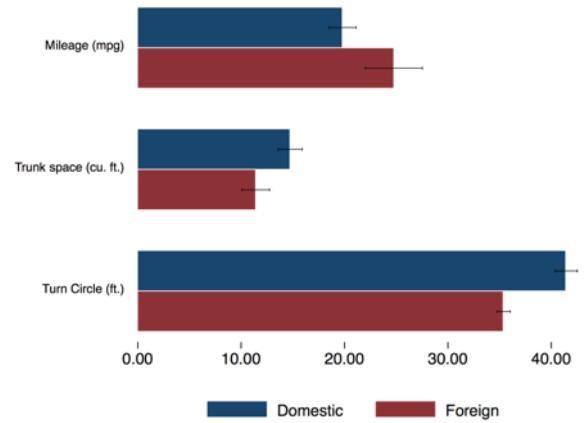
Stata Commands for Data Management and Analysis

View the Project on GitHub
worldbank/stata

Commands for Data Analysis

betterBar

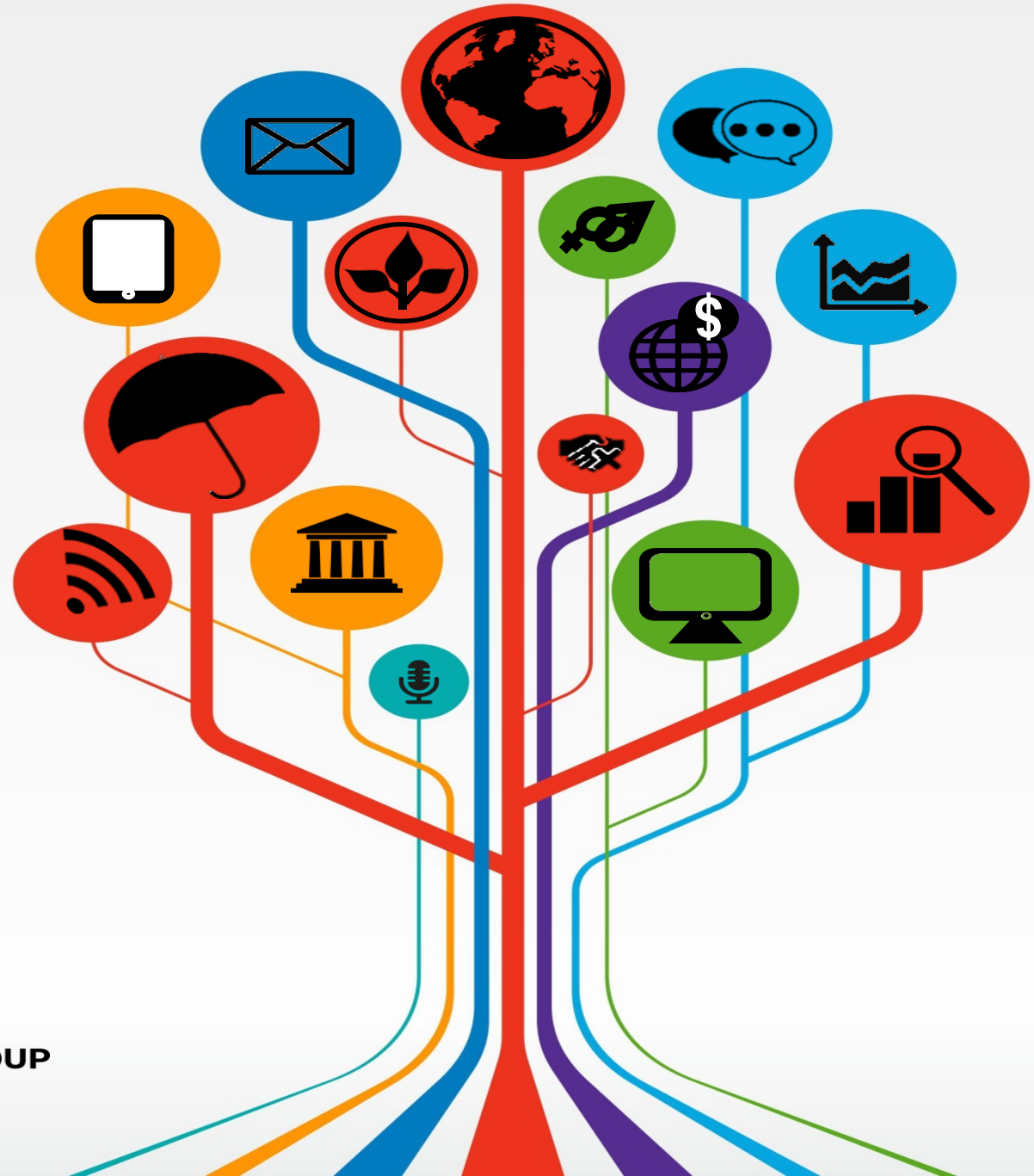
`betterBar` creates bar graphs for multiple variables with confidence intervals, setting `by()` and `over()` groups, adding labels and legends, and various styling commands.




```
wb_git_install betterBar
sysuse auto , clear
betterBar mpg trunk turn ///
  , over(foreign) se ///
  barlook(1 lw(thin) lc(white) fi(100))
```

Thank you!

Mrijan Rimal & Kristoffer Bjarkefur



i2i DIME
TRANSFORM DEVELOPMENT

WORLD BANK GROUP