Expanding your graphical repertoire

Richard Layton

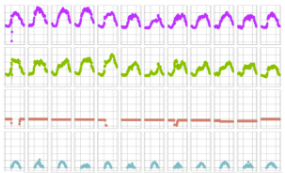
2024-02-13

Richard Layton resides online at

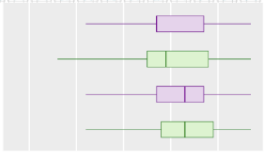
- `https://www.graphdoctor.com`
- `https://github.com/graphdr`

Variables, design, message

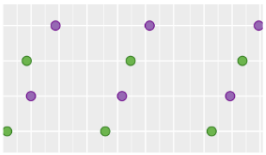
Trees, Maps, and Theorems by Jean-luc Doumont (2009) inspired the four main topics.



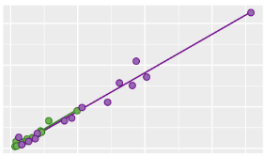
Showing evolution



Displaying distributions



Comparing data

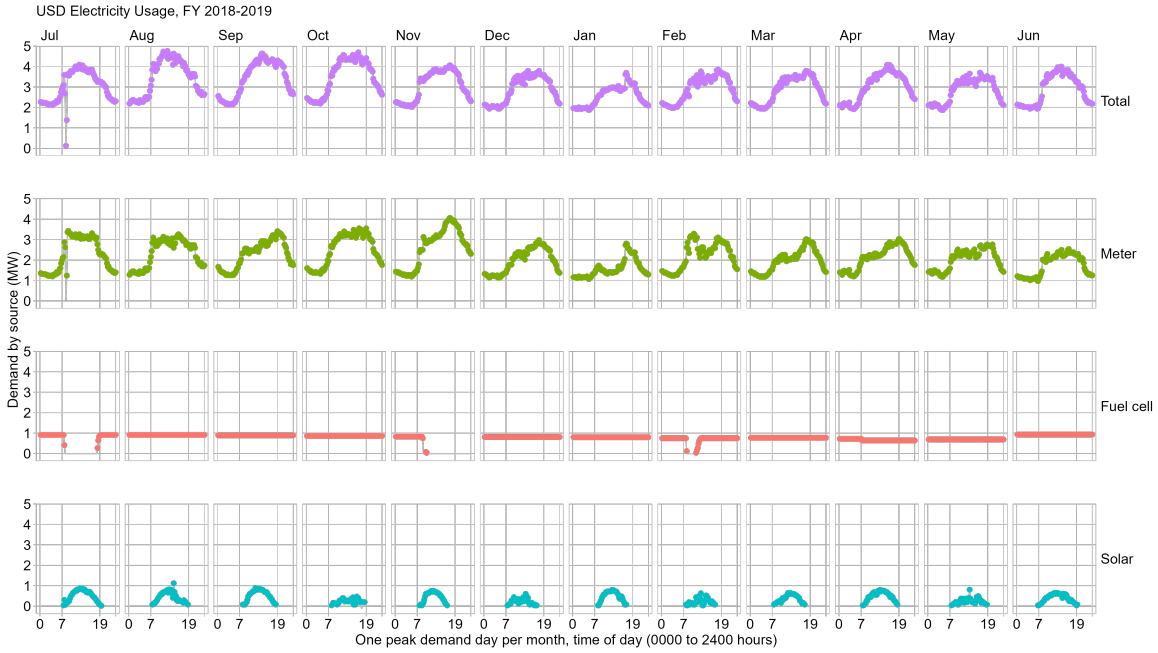


Revealing correlations

§ *Showing evolution*

Square brackets *[i]* give the slide number.

[4] Time series



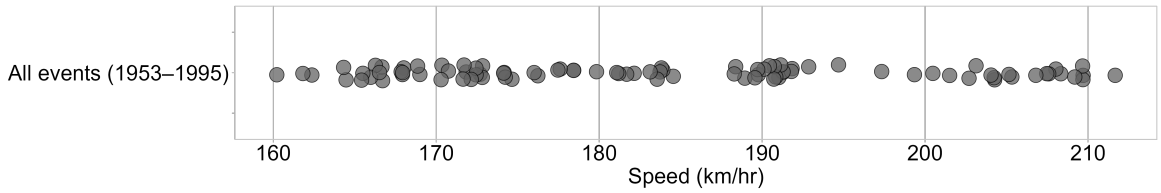
§ *Displaying distributions*

[6] Data

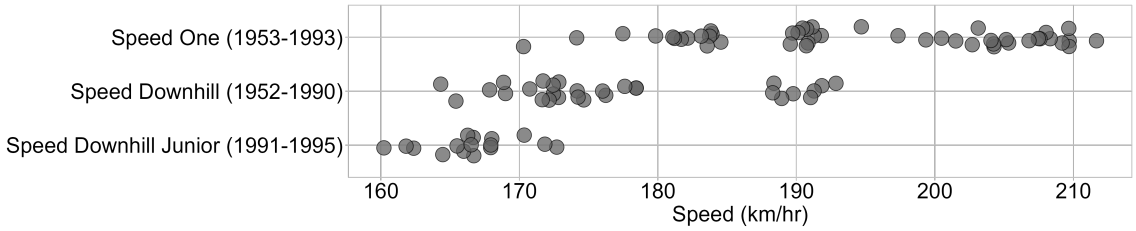
World speed skiing (km/hr) competitions 1953–1995

```
     Year                 Event    Sex  Speed
    <int>                <fctr> <fctr>  <num>
 1:  1952        Speed Downhill   Male 167.85
 2:  1953        Speed Downhill   Male 168.86
 3:  1953             Speed One   Male 174.14
 4:  1957             Speed One   Male 177.47
 5:  1961        Speed Downhill   Male 165.42
 6:  1961             Speed One   Male 190.12
---
86:  1993             Speed One   Male 170.30
87:  1994 Speed Downhill Junior Female 160.22
88:  1994 Speed Downhill Junior   Male 164.47
89:  1995 Speed Downhill Junior Female 166.52
90:  1995 Speed Downhill Junior Female 162.37
91:  1995 Speed Downhill Junior   Male 168.01
```

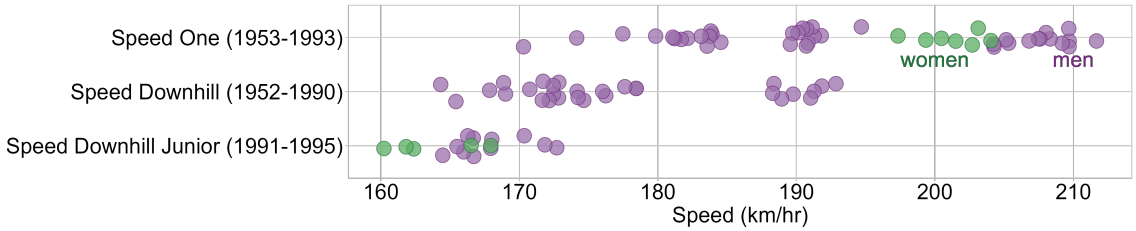
[7] Strip chart



[8] Add a category



[9] Add a second category

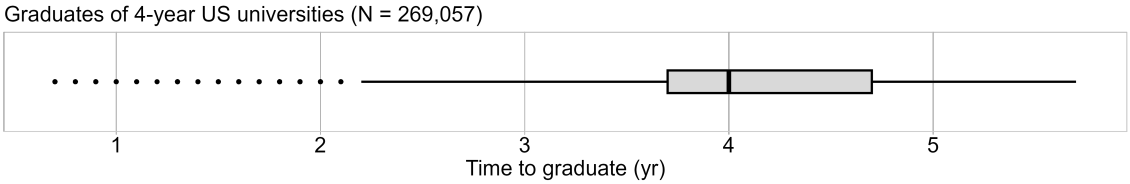


[10] Data

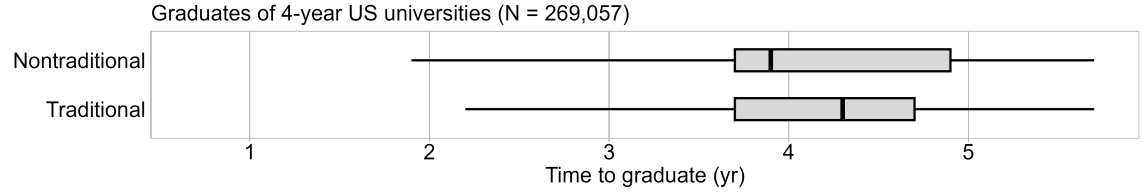
MIDFIELD graduates (N = 270k), enrolled in Engineering, excluding 10th and 90th quantiles

```
                path    sex years_to_grad
              <char> <char>         <num>
     1: Nontraditional Female           3.9
     2: Nontraditional Female           1.9
     3: Nontraditional Female           3.9
     4: Nontraditional Female           5.3
     5: Nontraditional Female           5.1
    ---
269053:    Traditional   Male           2.6
269054:    Traditional   Male           1.3
269055:    Traditional   Male           3.0
269056:    Traditional   Male           5.3
269057:    Traditional   Male           0.7
```

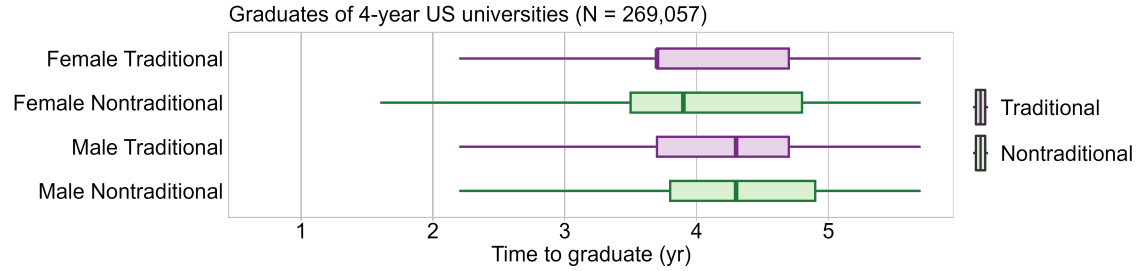
[11] Box and whisker chart



[12] Add a category



[13] Combine a second category

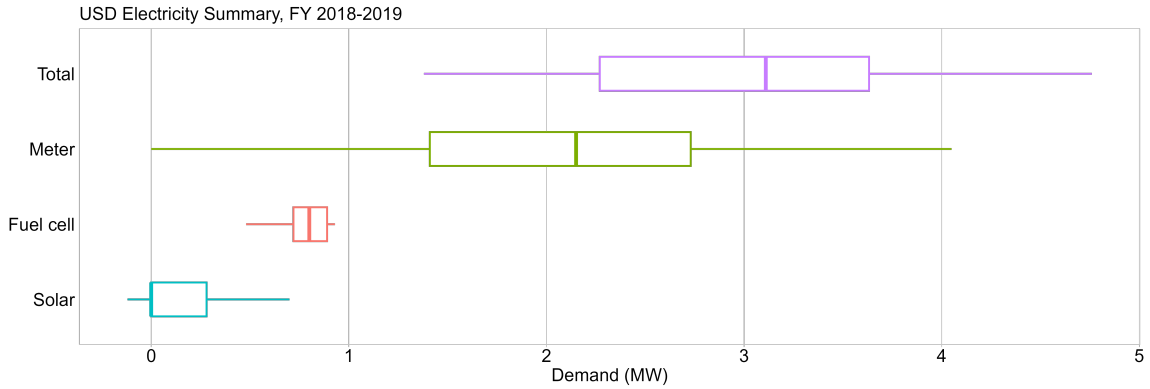


[14] Data

Revisiting the USD electrical data

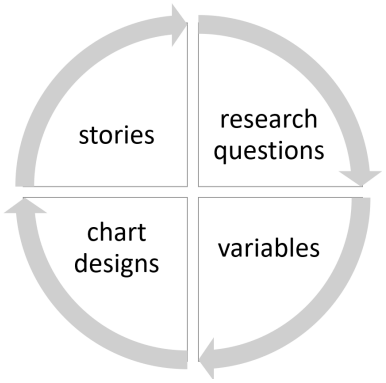
```
      bill_year bill_month date_peak minutes    source    MW
          <int>     <fctr>    <char>   <num>    <char> <num>
   1:      2019        Jan 1/22/2019      15 Fuel cell  0.80
   2:      2019        Jan 1/22/2019      15     Meter  1.16
   3:      2019        Jan 1/22/2019      15     Solar  0.00
   4:      2019        Jan 1/22/2019      15     Total  1.96
   5:      2019        Jan 1/22/2019      30 Fuel cell  0.80
  ---
4604:      2018        Sep 9/13/2018    1425     Total  2.76
4605:      2018        Sep 9/13/2018    1440 Fuel cell  0.89
4606:      2018        Sep 9/13/2018    1440     Meter  1.75
4607:      2018        Sep 9/13/2018    1440     Solar  0.00
4608:      2018        Sep 9/13/2018    1440     Total  2.64
```

[15] Ignoring time



[16] Discussion: Displaying distributions

Quantitative test scores from a recent exam could be displayed as a distribution. What categorical variable(s) could be added to create comparative distributions?



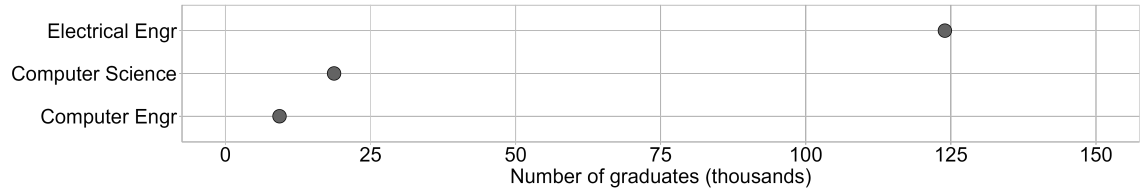
§ Comparing data

[18] Data

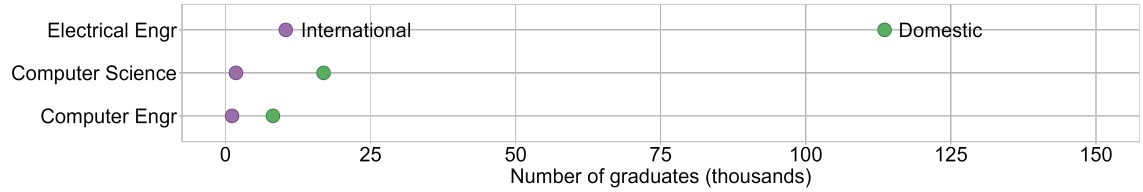
Representation at graduation in 3 engineering programs, 19 US institutions, 1987–2018

origin	sex	Electrical Engr	Computer Engr	Computer Science
<char>	<char>	<int>	<int>	<int>
1: International	Female	1865	140	365
2: International	Male	8530	993	1442
3: Domestic	Female	23426	702	2923
4: Domestic	Male	90150	7481	13987

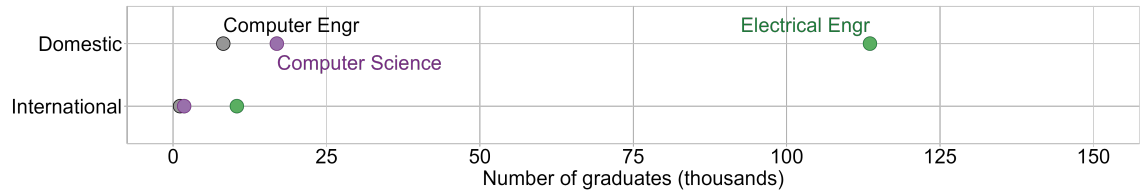
[19] Dot chart



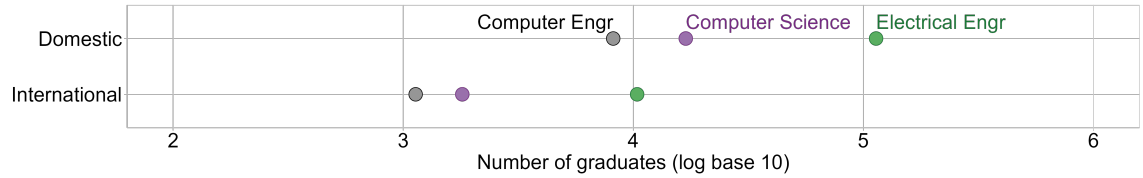
[20] Add a second category



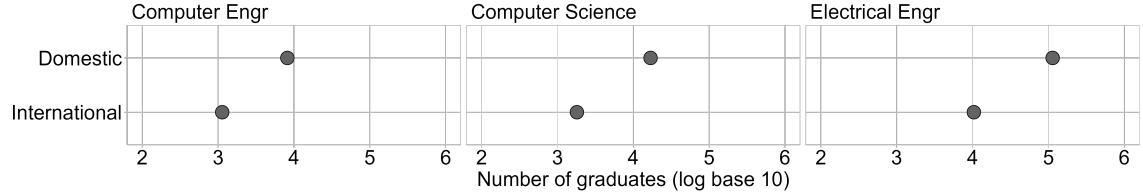
[21] Exchange mapping of categorical variables



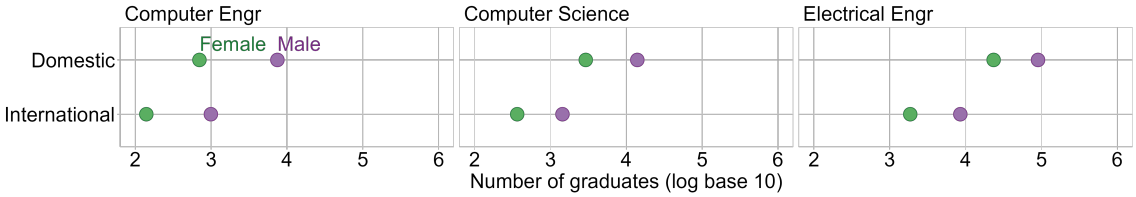
[22] Logarithmic scale for orders of magnitude differences



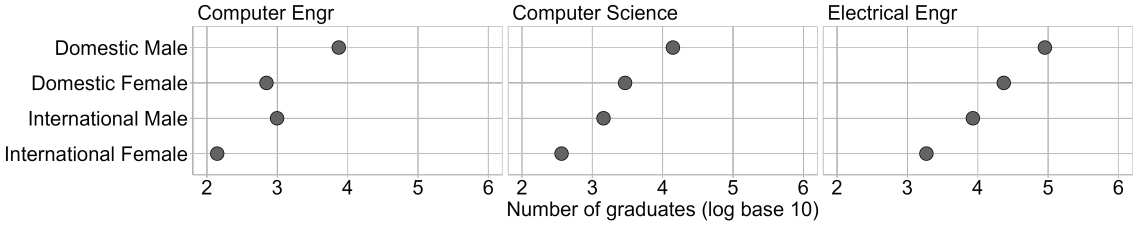
[23] One program per facet



[24] Add a third category

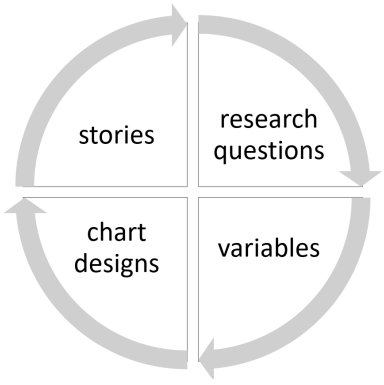


[25] Combine categories



[26] Discussion: Comparing data

Consider Table 2 Campus Buildings in the USD Energy Master Plan (p. 17-18). If we were to visualize these data in dot-chart form:



- Select the quantitative variable

- Select a categorical variable for the rows

- How would you order the rows?

- Select a second categorical variable for the facets

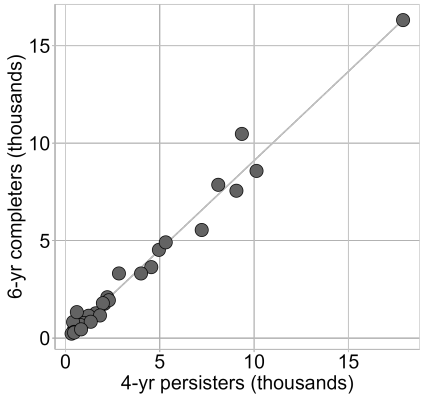
§ Revealing correlations

[28] Data

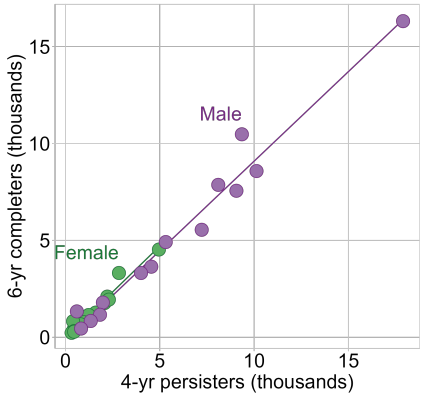
Engineering students at 14 institutions persisting to year 4 and graduating by year 6, 1987-2019

```
    institution    sex    y4    y6
         <char> <char> <int> <int>
 1:           A Female  4953  4525
 2:           A   Male 17897 16312
 3:           B Female  2834  3316
---
26:           N   Male  1338   838
27:           P Female   457   283
28:           P   Male   827   447
```

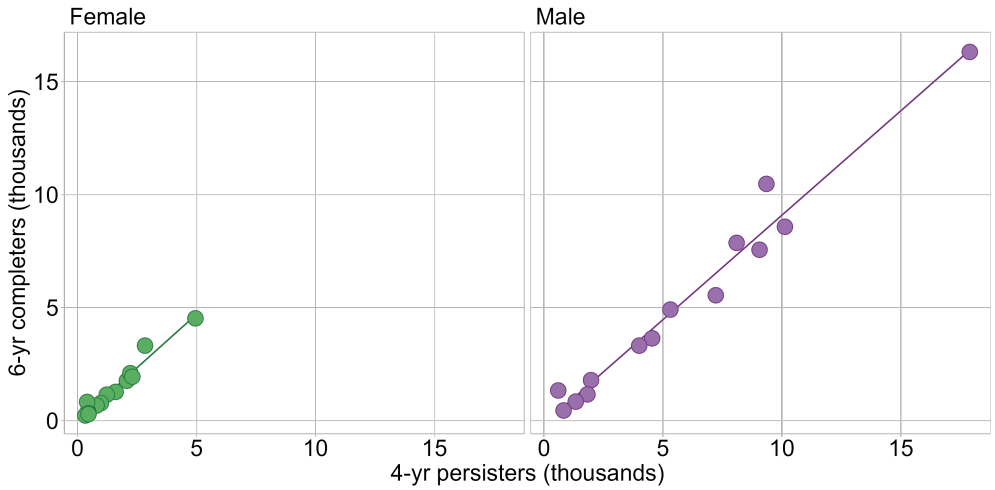
[29] Scatterplots are designed to reveal correlation



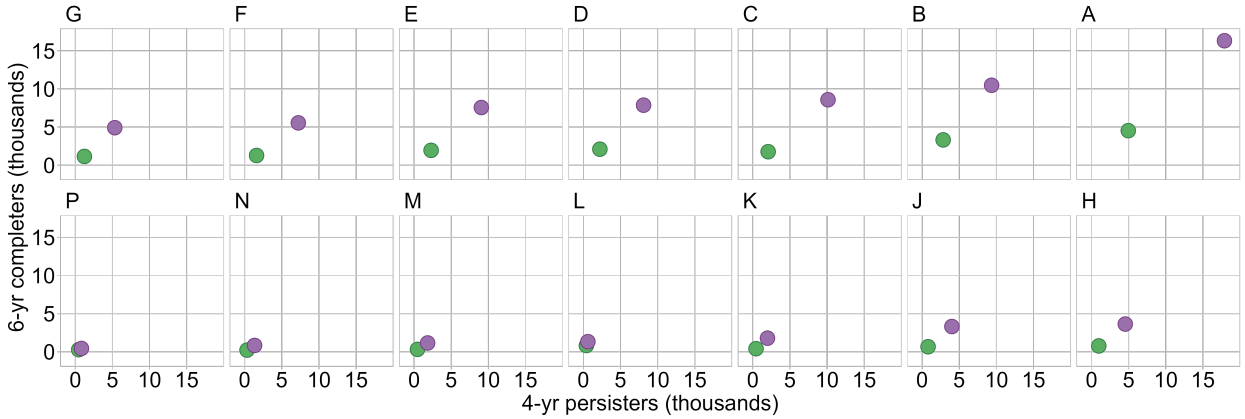
[30] Add a category



[31] One facet per sex

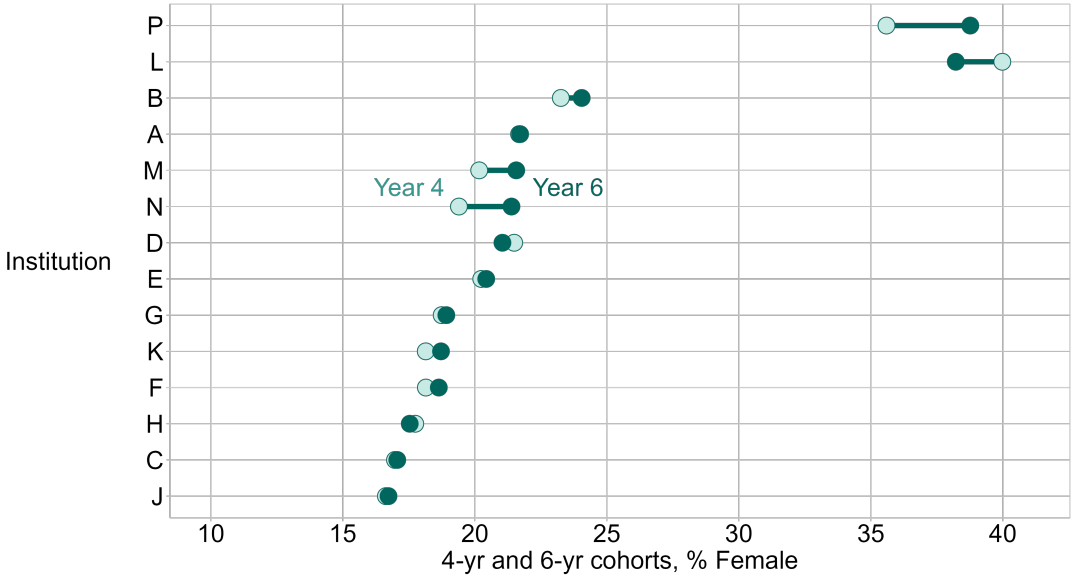


[32] One facet per institution



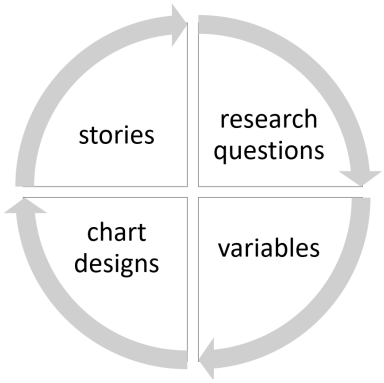
[33] Change the quantitative variable

Engineering students at 14 institutions persisting to year 4 and graduating by year 6, 1987–2019



[34] Discussion: Revealing correlations

Figure 17 (p. 36) of the USD Energy Master Plan has the form of a scatterplot. We must assume that the authors are attempting to discover a correlation between two quantitative variables.



- What are the two quantitative variables?

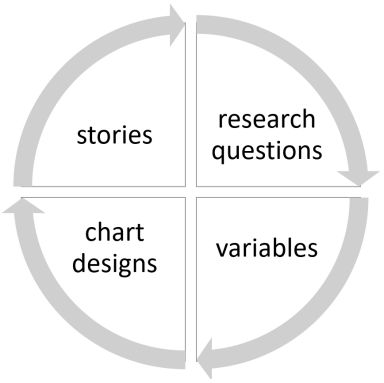
- Do the variables appear to be correlated?

- Is the linear curve fit justified?

§ *Closing discussion*

[36] Variables, design, message

- Chart design depends on your variables
- Chart design depends on the message the data convey
- Continue to expand your repertoire of chart types



References

Clarke, Steven, Michael Anderson, Daniela Aramayo, Dominic Molinari, Arthur Tseng, Zoe Warp, and John Ko. 2021. "University of San Diego Energy Master Plan." Anaheim, CA: Willdan Energy Solutions.

Doumont, Jean-luc. 2009. *Trees, Maps, and Theorems*. Belgium: Principiae.

Unwin, Antony. 2015. *GDAdata: Datasets for the Book Graphical Data Analysis with r*. `https://CRAN.R-project.org/package=GDAdata`.