# *Expanding your graphical repertoire*

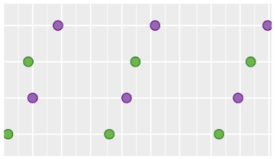
## *2024 MIDFIELD Institute*

*Richard Layton* resides online at

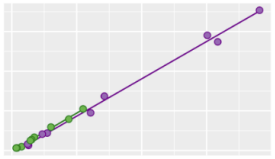
- `https://www.graphdoctor.com`
- `https://github.com/graphdr`

## *Variables, design, message*

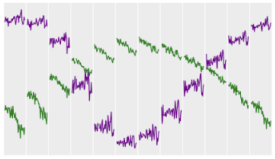
*Trees, Maps, and Theorems* by Jean-luc Doumont (2009) inspired the four main topics.



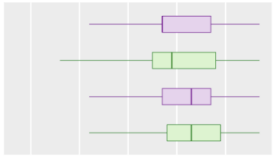
**Comparing data**



**Revealing correlations**



**Showing evolution**



**Displaying distributions**

## § Comparing data

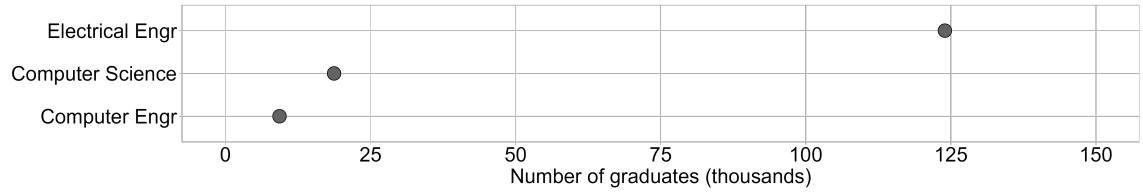
### [4] Data

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

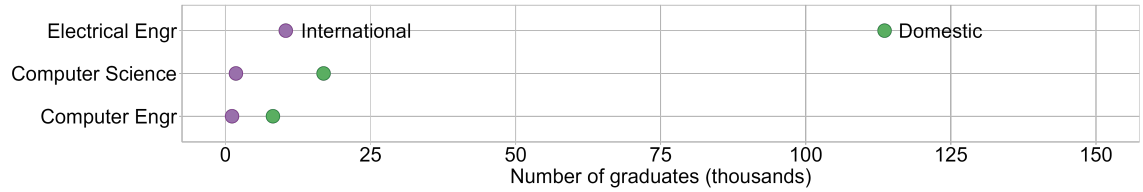
Representation at graduation in 3 engineering programs, 19 US institutions, 1987–2018

| | 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 |

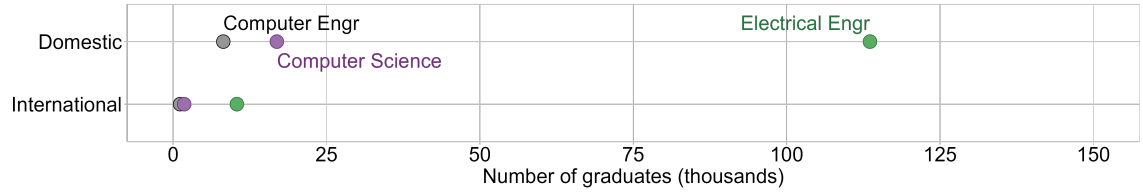
### [5] Dot chart



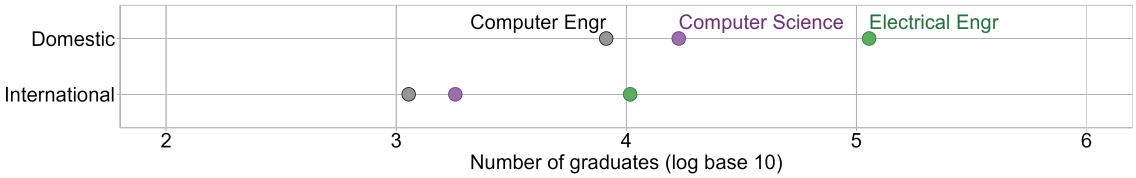

### [6] Add a second category



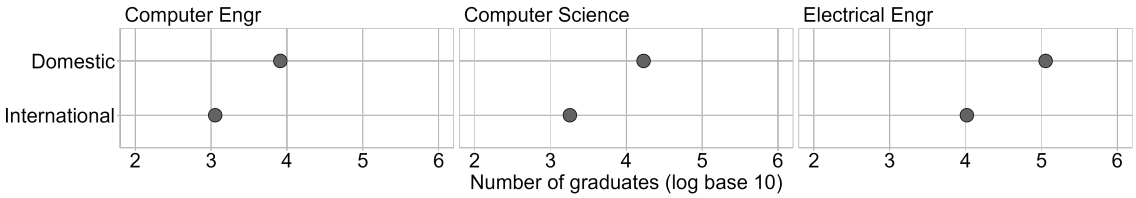

### [7] Exchange mapping of categorical variables



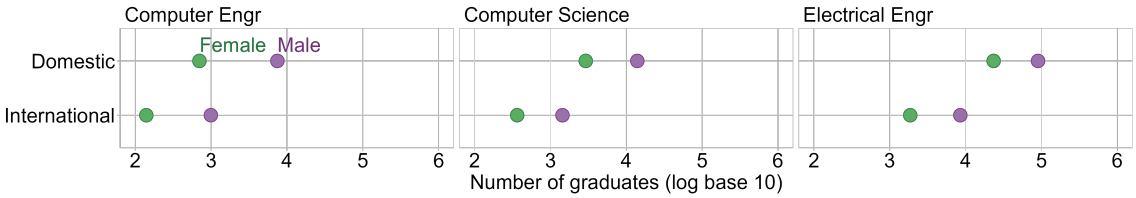
*[8] Logarithmic scale for orders of magnitude differences*



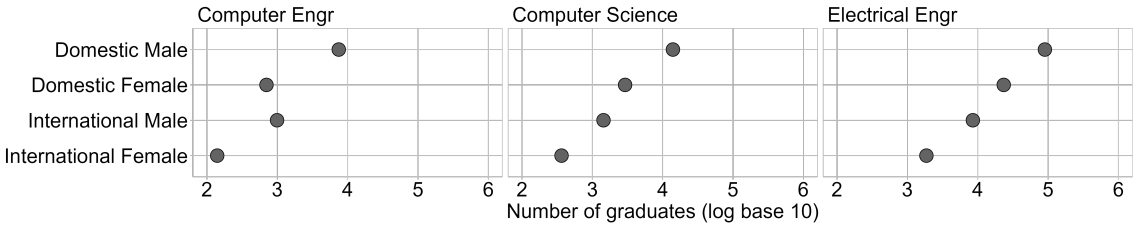

*[9] One program per facet*




*[10] Add a third category*

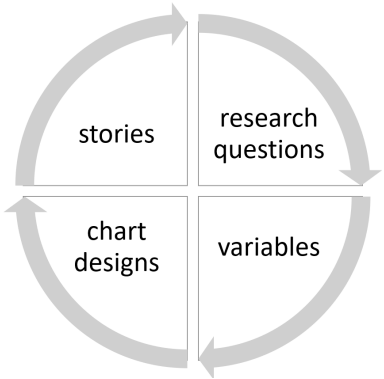



*[11] Combine categories*




*[12] Discussion: Comparing data*

What points seem most important to you so far?



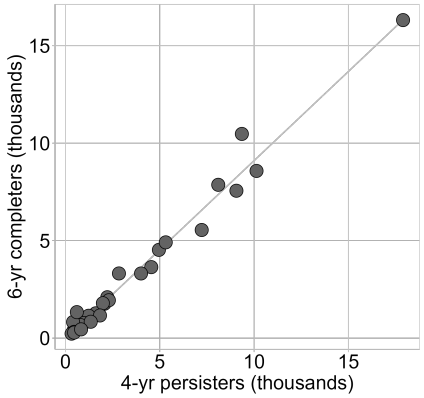
## § *Revealing correlations*

### *[14] Data*

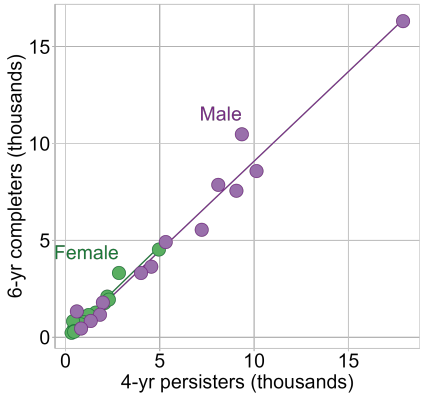
Engineering students at 14 institutions persisting to year 4 and graduating by year 6, 1987-2019

```
    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
```

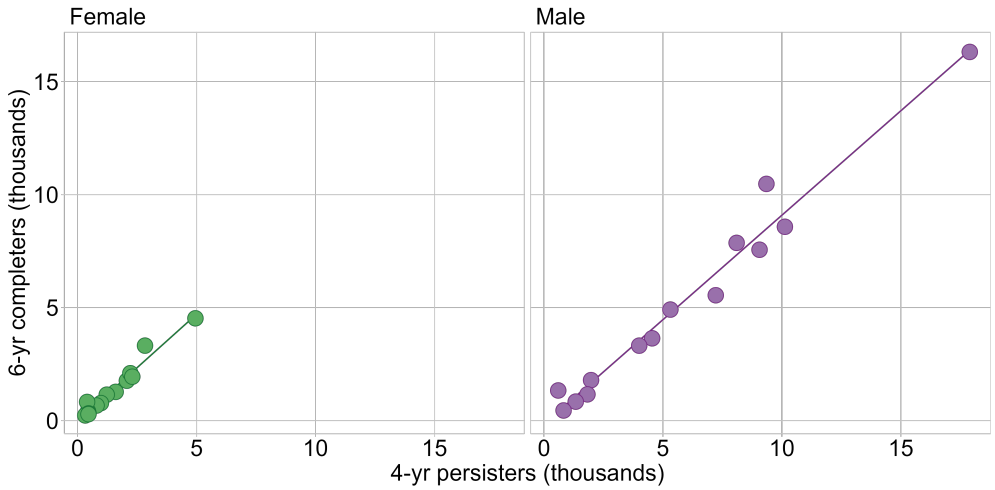
### *[15] Scatterplots are designed to reveal correlation*



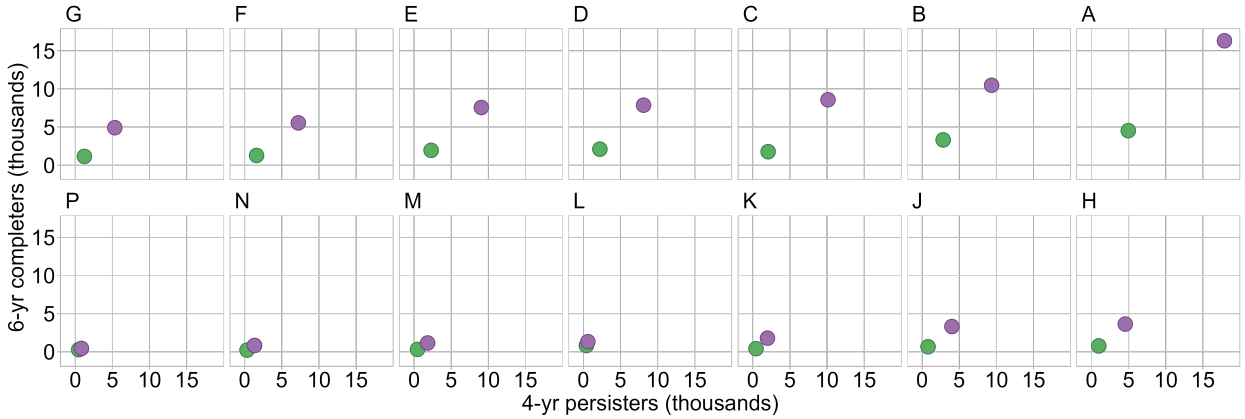

### *[16] Add a category*



*[17] One facet per sex*

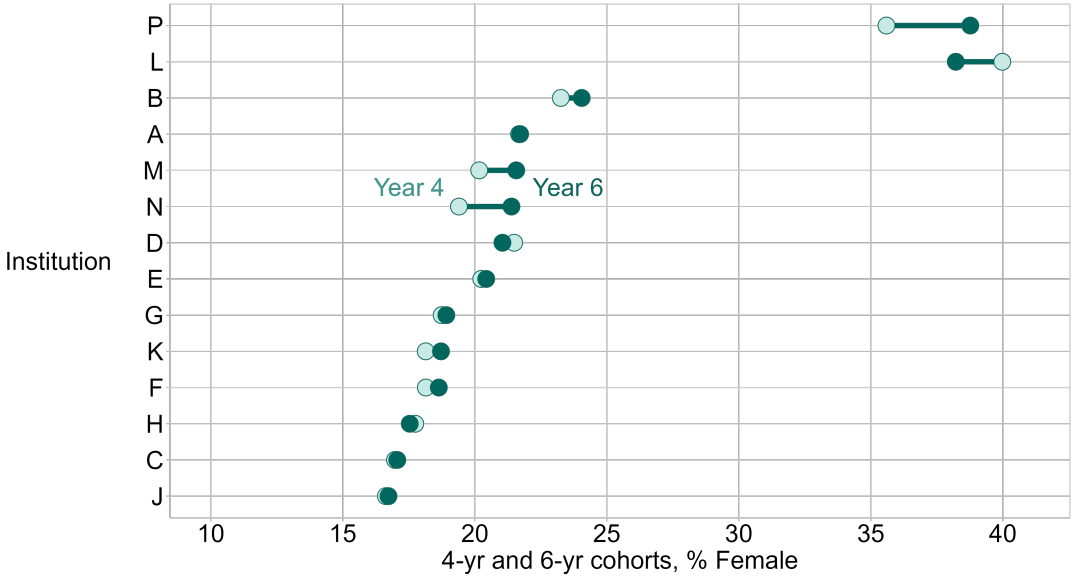



*[18] One facet per institution*



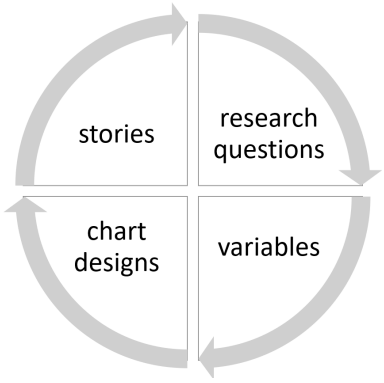
*[19] Change the quantitative variable*

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




*[20] Discussion: Revealing correlations*

- We saw a correlation.
- We changed the emphasis.
- Which chart tells a more compelling story?



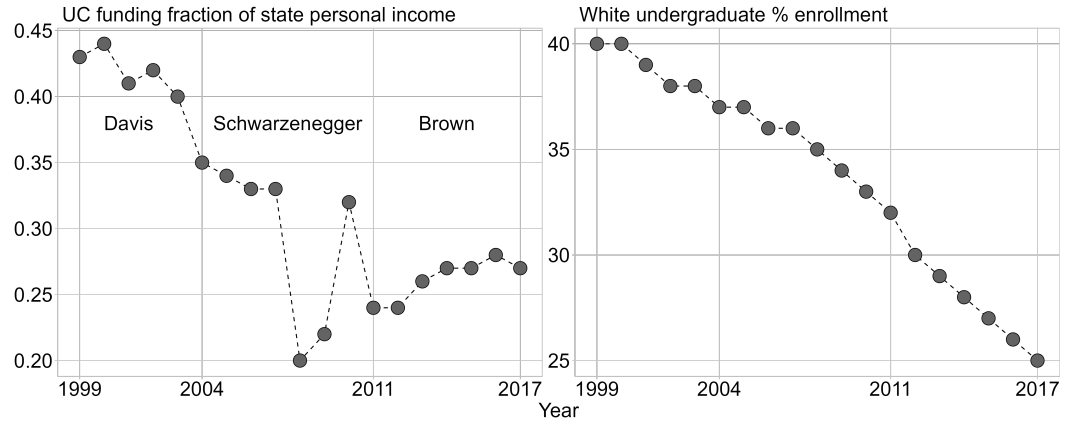
## § *Showing evolution*

### *[22] Data*

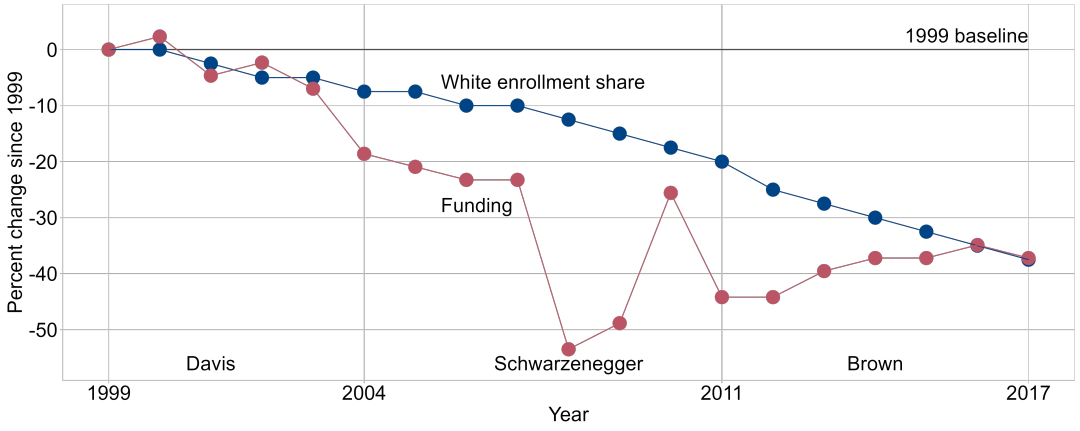
University of California: funding and percent White enrollment, 1999–2017

```
     Year Governor Pct_UG_White Funding_metric
    <num>   <char>        <num>          <num>
 1:  1999    Davis           40           0.43
 2:  2000    Davis           40           0.44
 3:  2001    Davis           39           0.41
---
17:  2015    Brown           27           0.27
18:  2016    Brown           26           0.28
19:  2017    Brown           25           0.27
```

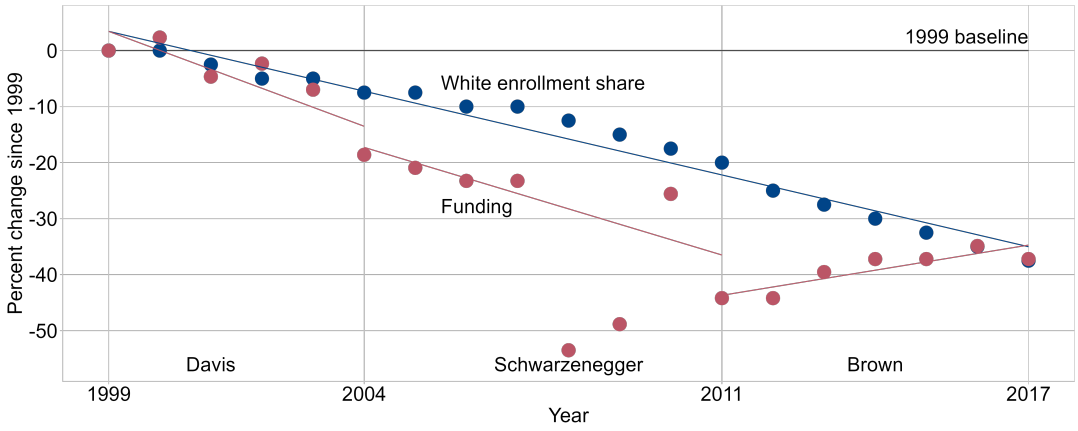
### *[23] Two time series*



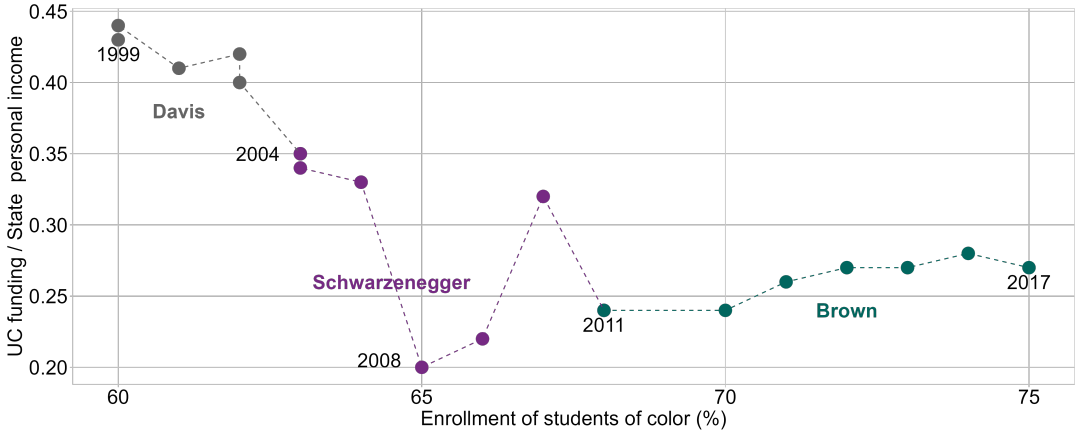

### *[24] Indexed time series*



*[25] Parallel lines indicate possible correlation*




*[26] Connected scatterplot*

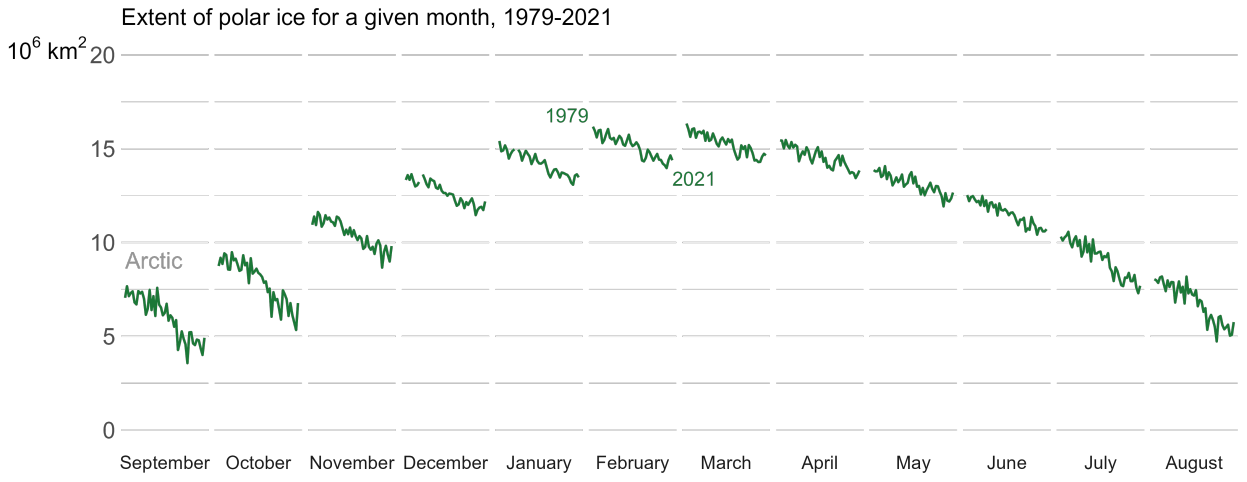



*[27] Data*

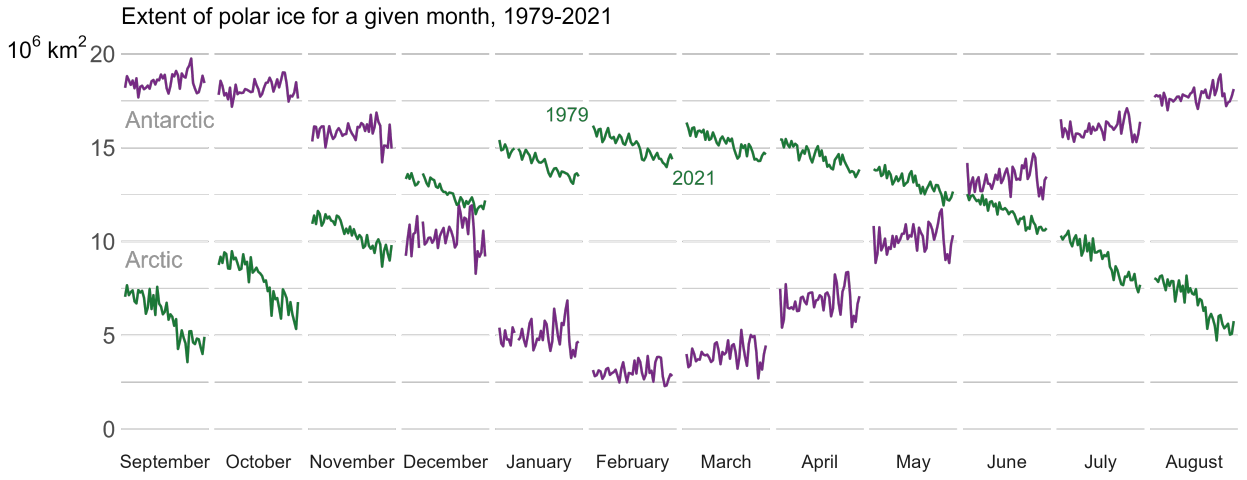
Extent of polar ice (millions sq km) 1979–2021

```
          hemis     month  year extent
         <char>    <fctr> <int>  <num>
   1:    Arctic September  1979  7.051
   2:    Arctic September  1980  7.667
   3:    Arctic September  1981  7.138
  ---
1030: Antarctic    August  2019 17.478
1031: Antarctic    August  2020 17.758
1032: Antarctic    August  2021 18.131
```

*[28] Cyclic time series*

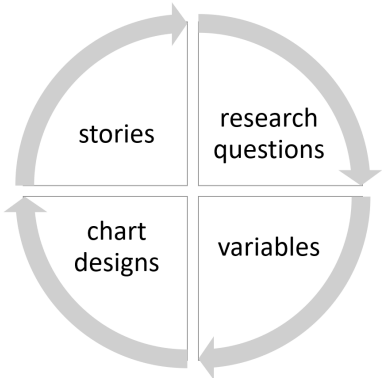



*[29] Add a category*




*[30] Discussion: Showing evolution*

- Which time series chart design might be used in your own work?
- Explain.



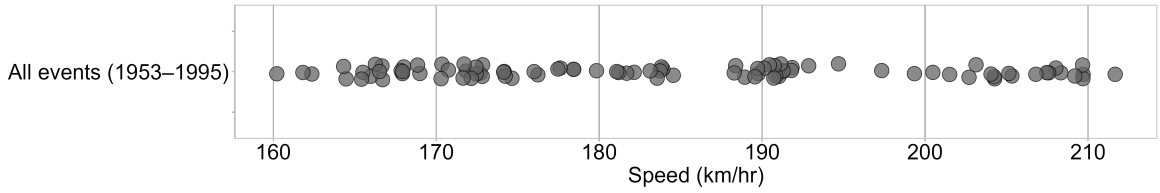
## § Displaying distributions

### *[32] Data*

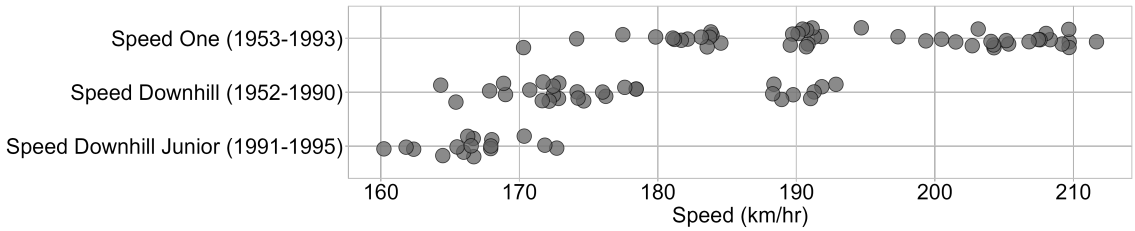
World speed skiing (km/hr) competitions 1953–1995

```
Key: <Event, Year, Sex>
             Event  Year    Sex  Speed
            <fctr> <int> <fctr>  <num>
 1: Speed Downhill  1952   Male 167.85
 2: Speed Downhill  1953   Male 168.86
 3: Speed Downhill  1961   Male 165.42
 4: Speed Downhill  1962   Male 172.85
---
88:      Speed One  1990 Female 199.35
89:      Speed One  1991   Male 207.59
90:      Speed One  1993   Male 208.33
91:      Speed One  1993   Male 170.30
```

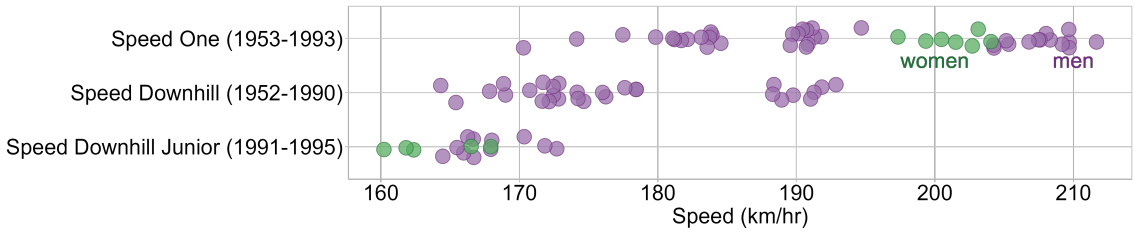
### *[33] Strip chart*




### *[34] Add a category*




### *[35] Add a second category*

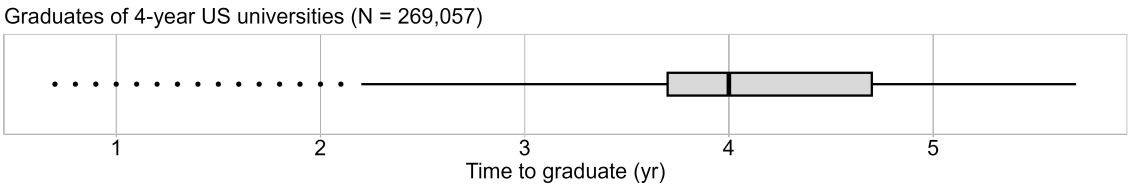


## *[36] Data*

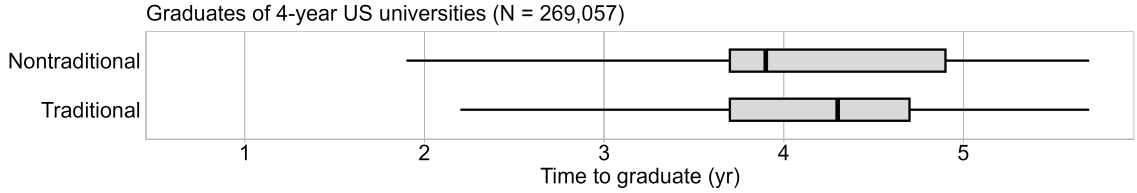
MIDFIELD graduates (N = 270k), enrolled in Engineering, excluding 10th and 90th quantiles

```
Key: <path, sex>
                 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
    ---
269054:    Traditional   Male           1.3
269055:    Traditional   Male           3.0
269056:    Traditional   Male           5.3
269057:    Traditional   Male           0.7
```

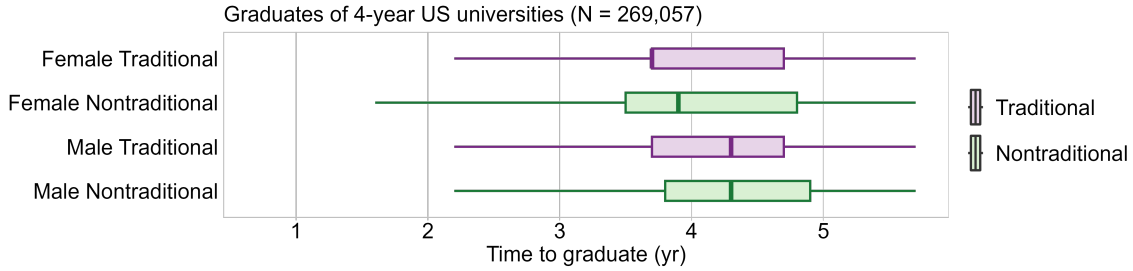
## *[37] Box and whisker chart*




## *[38] Add a category*



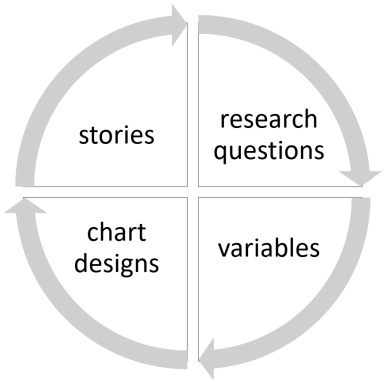

## *[39] Combine a second category*



*[40] Discussion: Displaying distributions*

What MIDFIELD distributions would you like to see:

- what quantitative variable?
- what categorical variables?




## § Closing discussion

*[42] Variables, design, message*

- For you, what was the muddiest point in the session?

- Is there a graph design you would have liked to have seen today?

- Is there a class of variables you would have liked to have seen today?

## *References*

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

F. Fetterer, K. Knowles, W.N. Meier, M. Savoie, and A.K. Windnagel. Sea ice index, version 3, Sea ice extent and area organized by year. 2017. DOI: https://doi.org/10.7265/N5K072F8. URL `https://nsidc.org/arcticseaicenews/sea-ice-tools/`.

Christopher Newfield. Budget justice: Addressing the structural racism of higher education funding. *Academe*, 107(2):57–64, 2021. URL `https://www.aaup.org/article/budget-justice`.

Antony Unwin. *GDAdata: Datasets for the Book Graphical Data Analysis with R*, 2015. URL `https://CRAN.R-project.org/package=GDAdata`. R package version 0.93.