

# Expanding your graphical repertoire

2019 MIDFIELD Institute

---

Richard Layton

June 2019

## Data structure determines which graph types are feasible



Number of observations?

Number of variables?  
Continuous or discrete?



Number of variables?  
Nominal or ordinal?  
Number of levels each?

## Graphical repertoire determines which graph types you explore

62

strip plot

box and whisker plot

multiway

scatterplot

dot plot

line graph

conditioning plot

scatterplot matrix

parallel coordinate plot

63

cycle plot

quantile-quantile plot

Sankey diagram

diverging stacked bar

multivariate bar

micromaps

proportional symbol map

dot density map

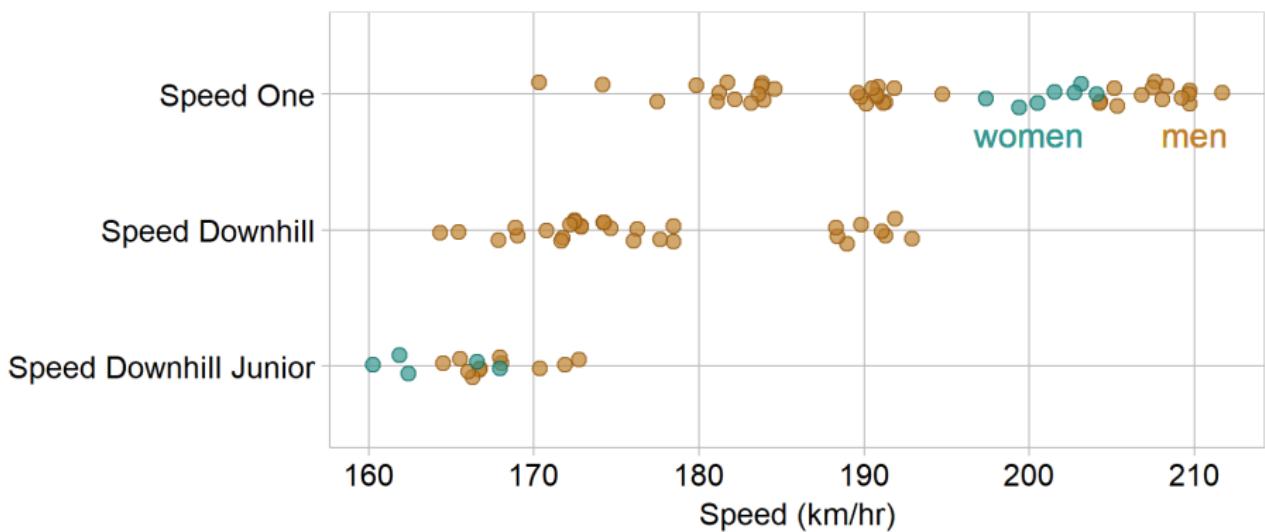
## Gallery of graph types

---

# Strip plot – or jitter plot or 1D scatterplot

Quantitative: Speed

Categorical: Event, sex

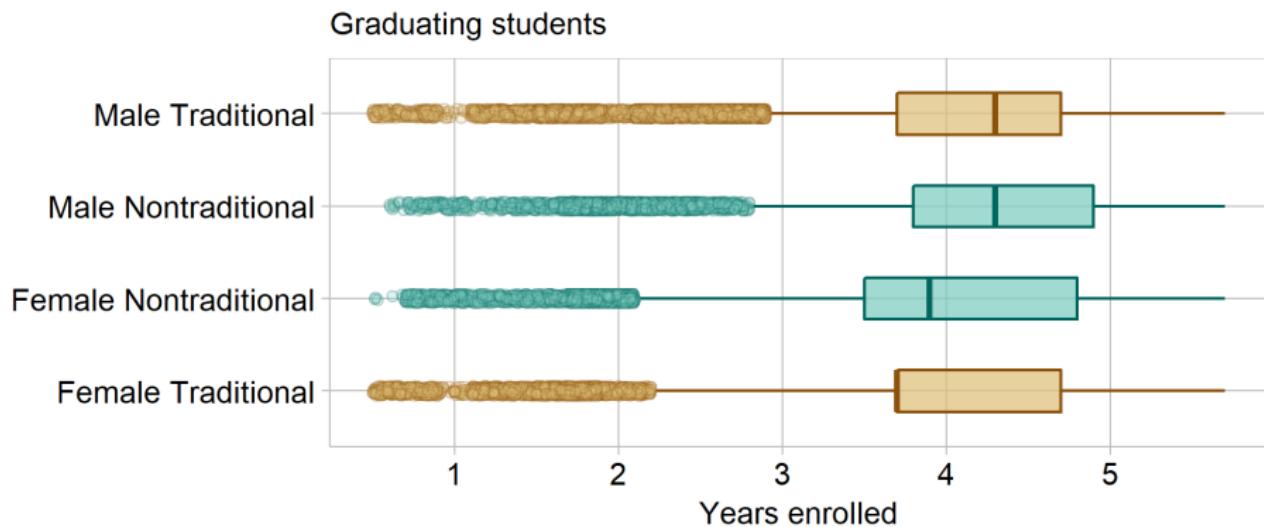


Data source (Unwin, 2015)

# Box and whisker plot or box plot

Quantitative: Years enrolled

Categorical: Path + sex

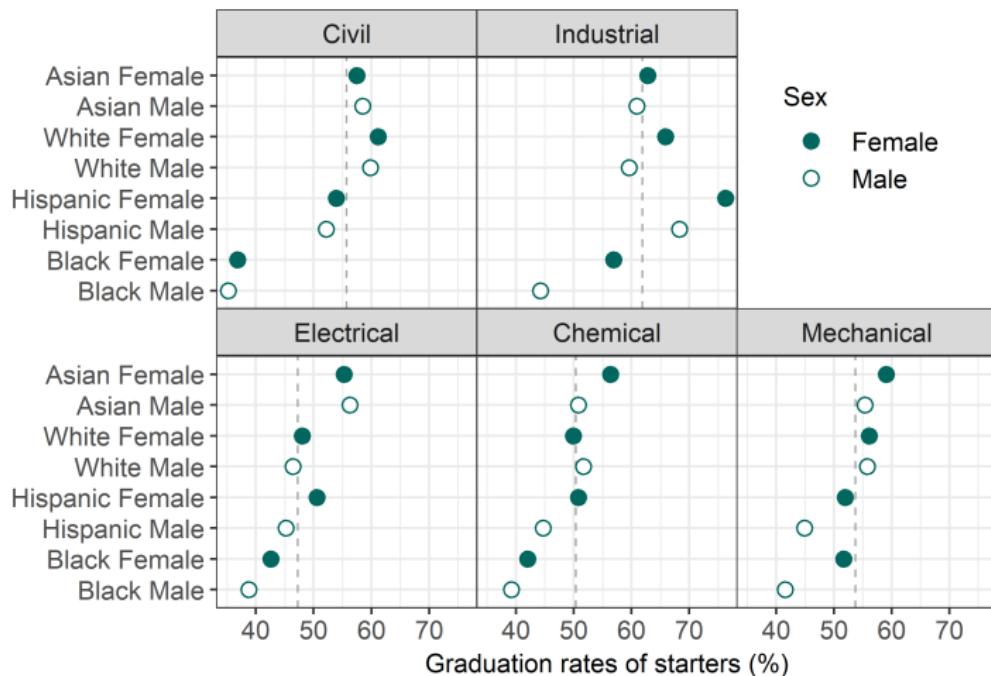


Data source (MIDFIELD, 2019)

# Multiway plot

Quantitative: Graduation rate

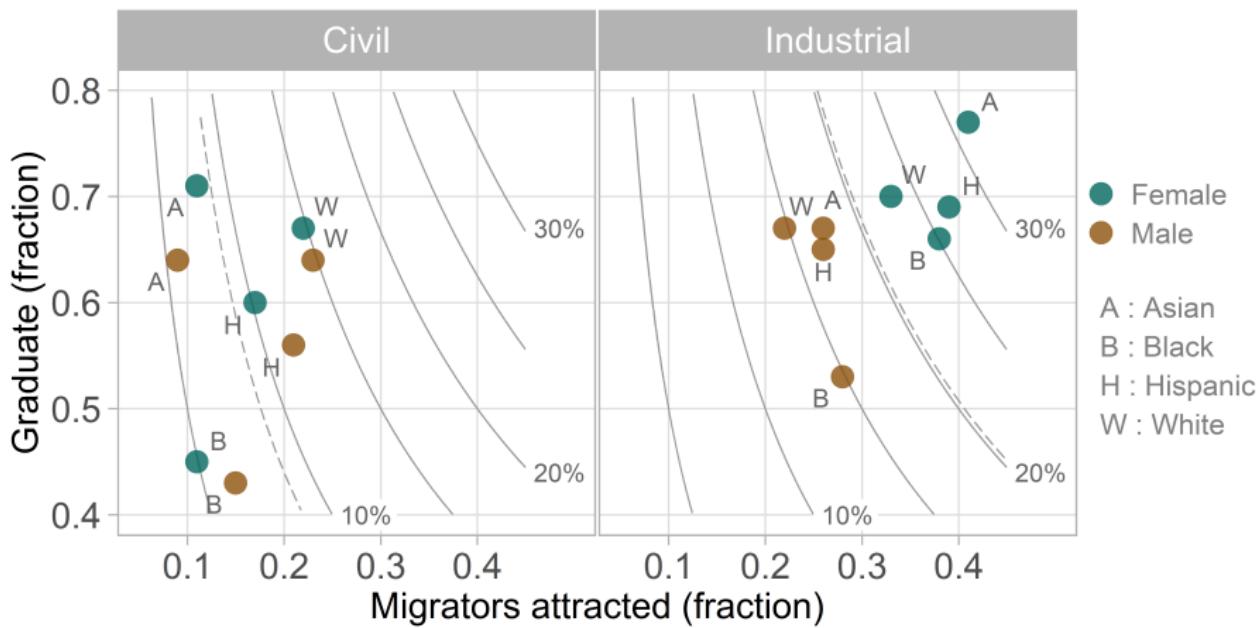
Categorical: Race/ethnicity/sex, major



# Scatterplot

Quantitative: Migrators attracted, migrants graduated

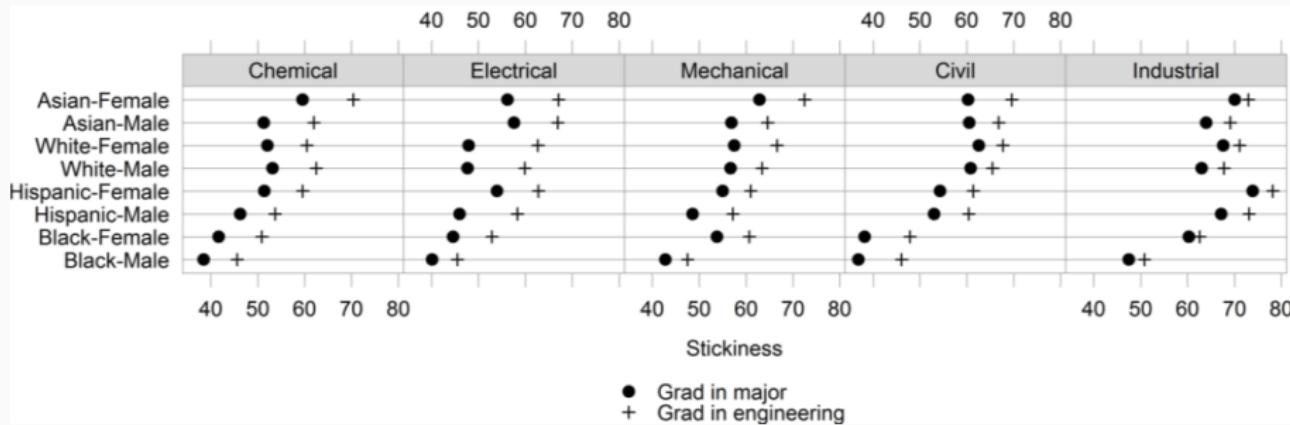
Categorical: Race/ethnicity, sex, major



# Cleveland dot plot

Quantitative: Stickiness

Categorical: Race/ethnicity/sex, major, graduation destination



Data source (MIDFIELD, 2019)

# Line graph

Death by Air Pollution in the United States Compared with the World

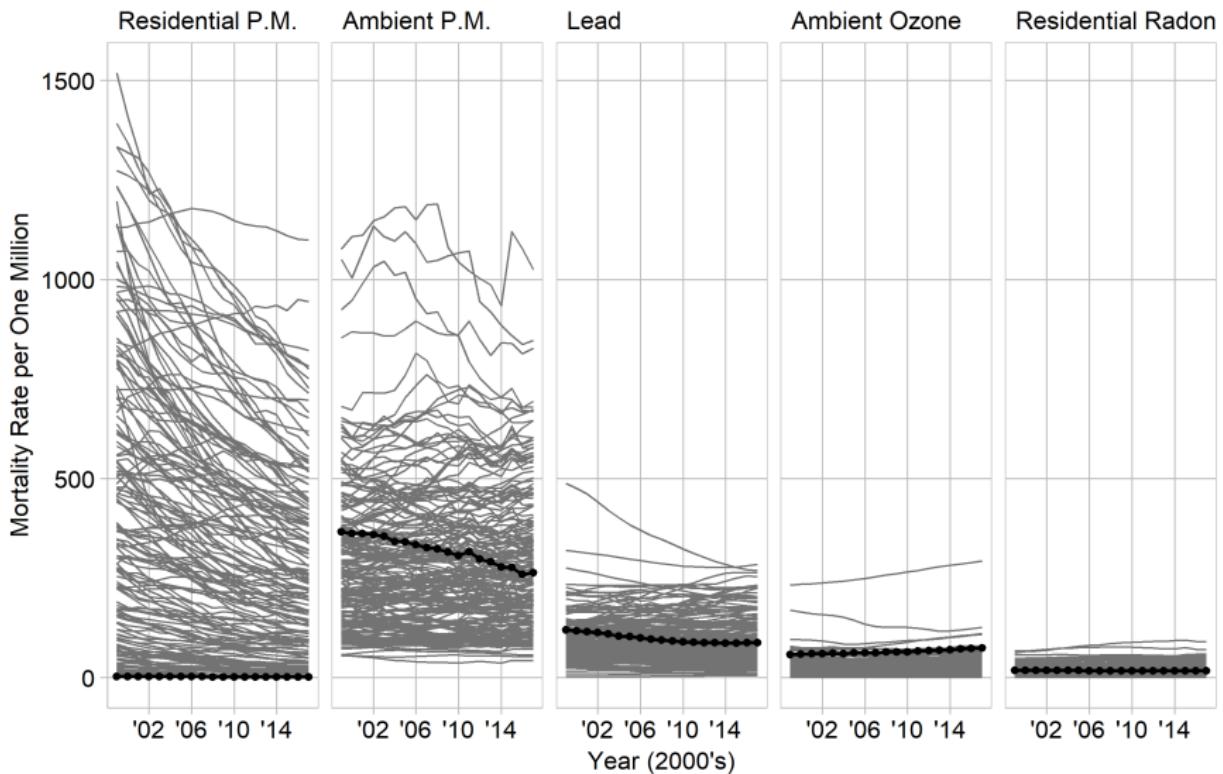
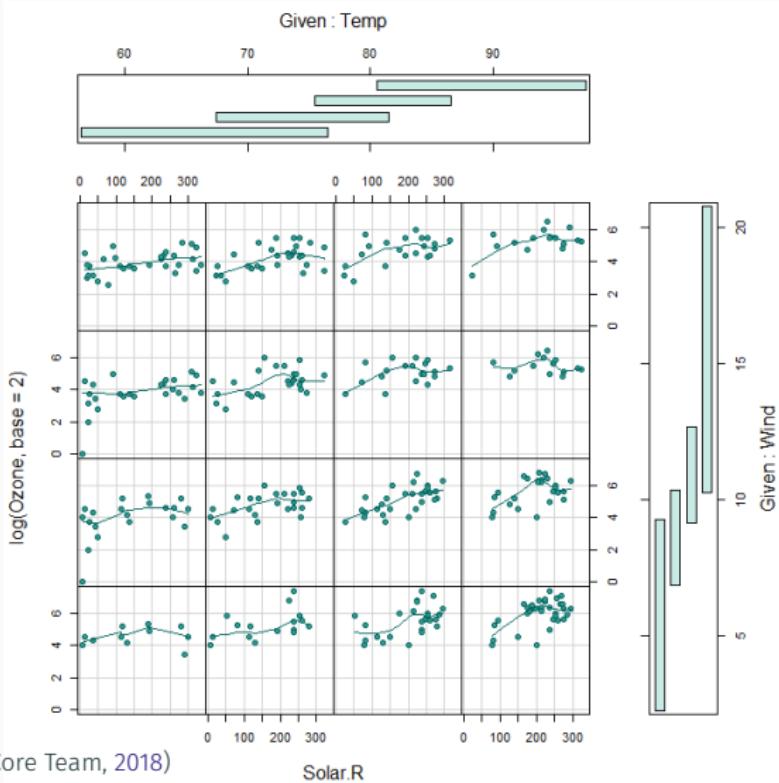


Image source (Ryan, 2019)

Source: OECD

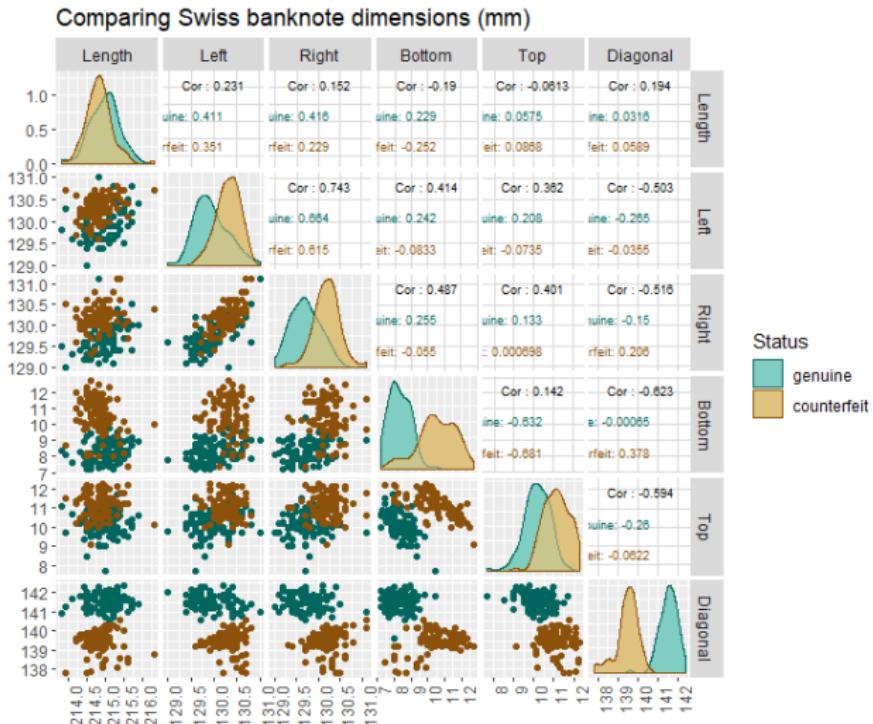
# Conditioning plot or co-plot

Quantitative: Ozone, temperature, wind speed, solar radiation



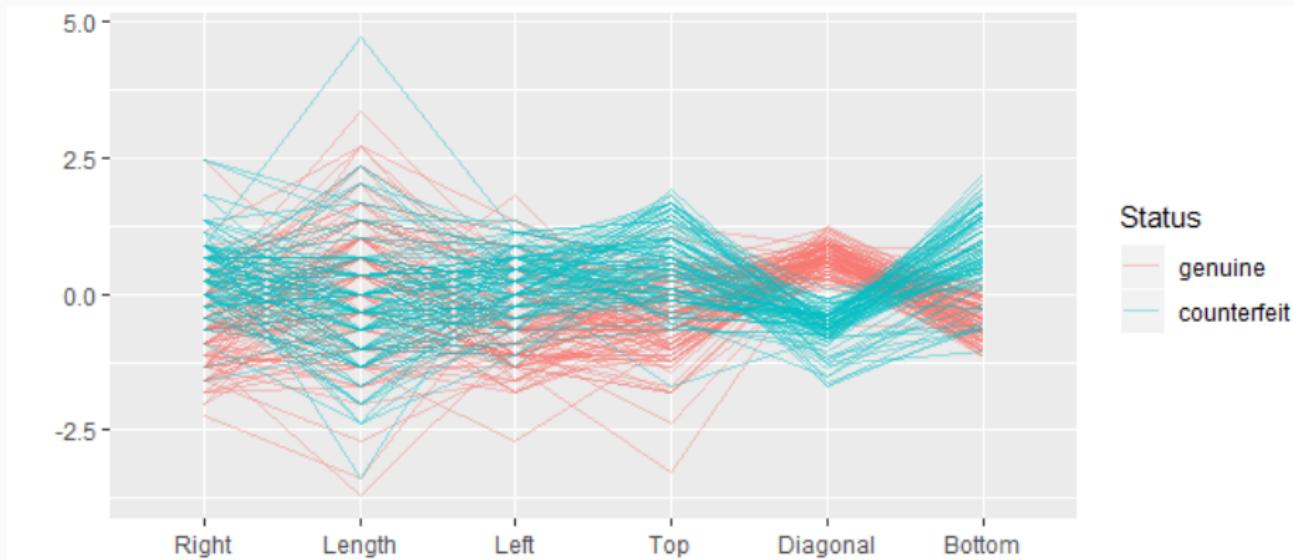
# Scatterplot matrix

Quantitative: Six dimensions of Swiss banknotes



# Parallel coordinate plot

Quantitative: Six dimensions of Swiss banknotes



# Cycle plot or month plot

Quantitative: Number of wildlife strikes on aircraft

Categorical: Month, year

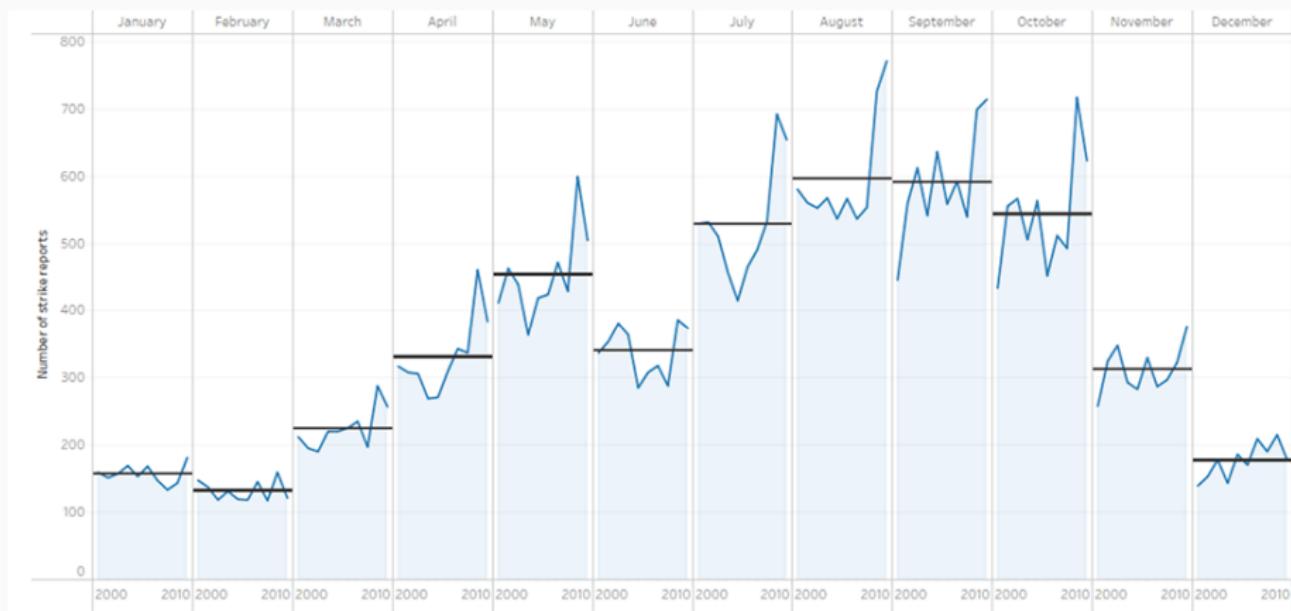
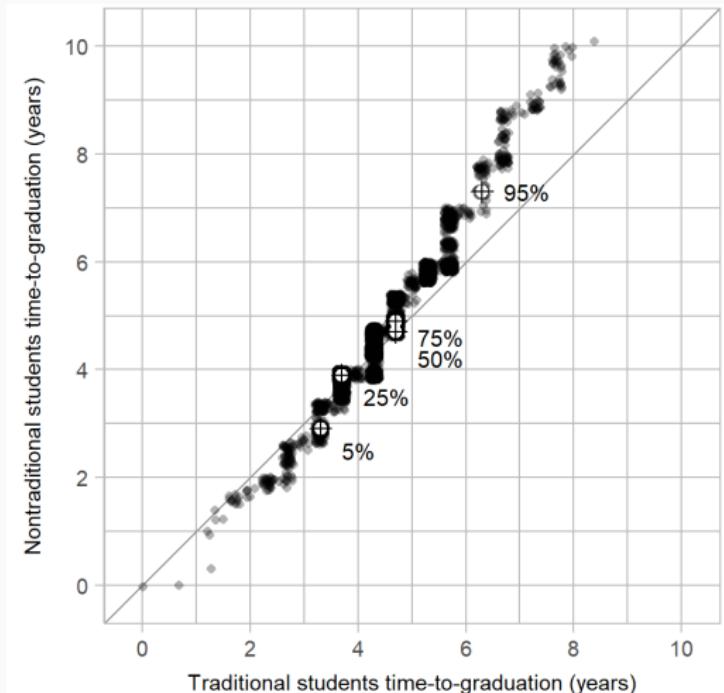


Image source <https://canonicalized.com/cycle-plots-tableau/>

# Quantile-quantile plot or q-q plot

Quantitative: Years to graduate quantiles

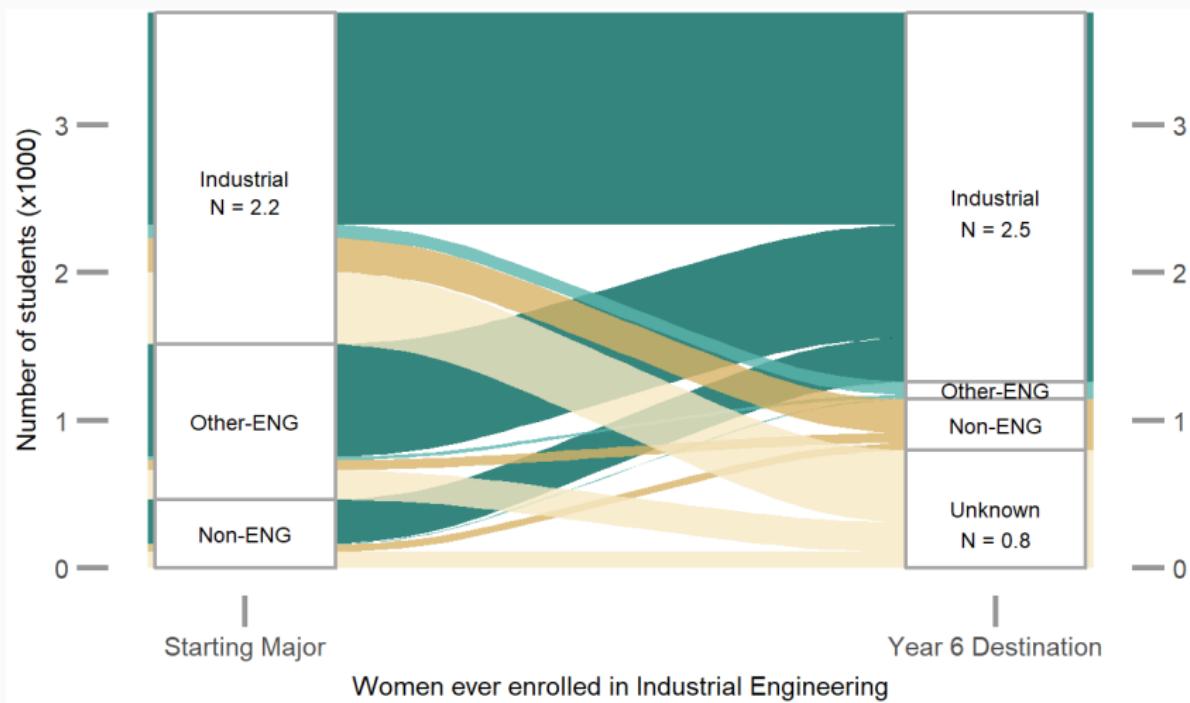
Categorical: Traditional students, nontraditional students



# Sankey diagram

Quantitative: Numbers of students

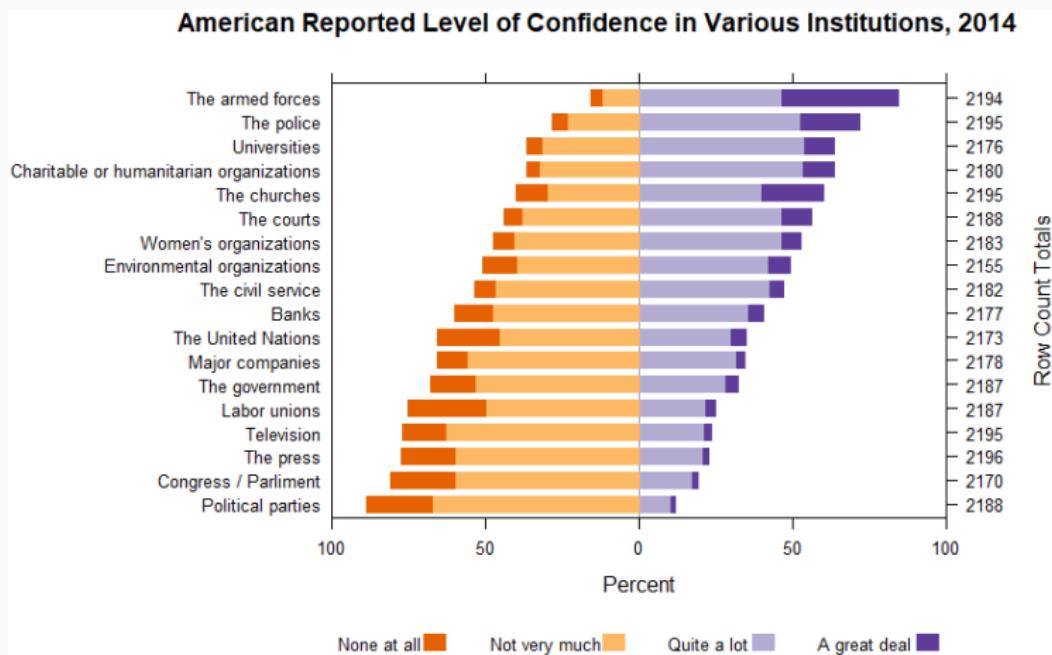
Categorical: Starting major, destination major



# Diverging stacked bar chart

Quantitative: Frequency of response, count totals

Categorical: Survey questions, level of confidence



# Multivariate bar chart

Quantitative

- number of people
- percent became paralyzed

Categorical

- age group
- vaccination status

Mantel-Haenszel-Cochran test of the Salk polio vaccine

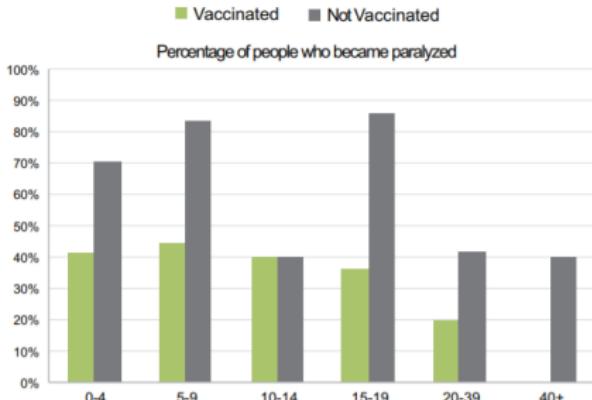
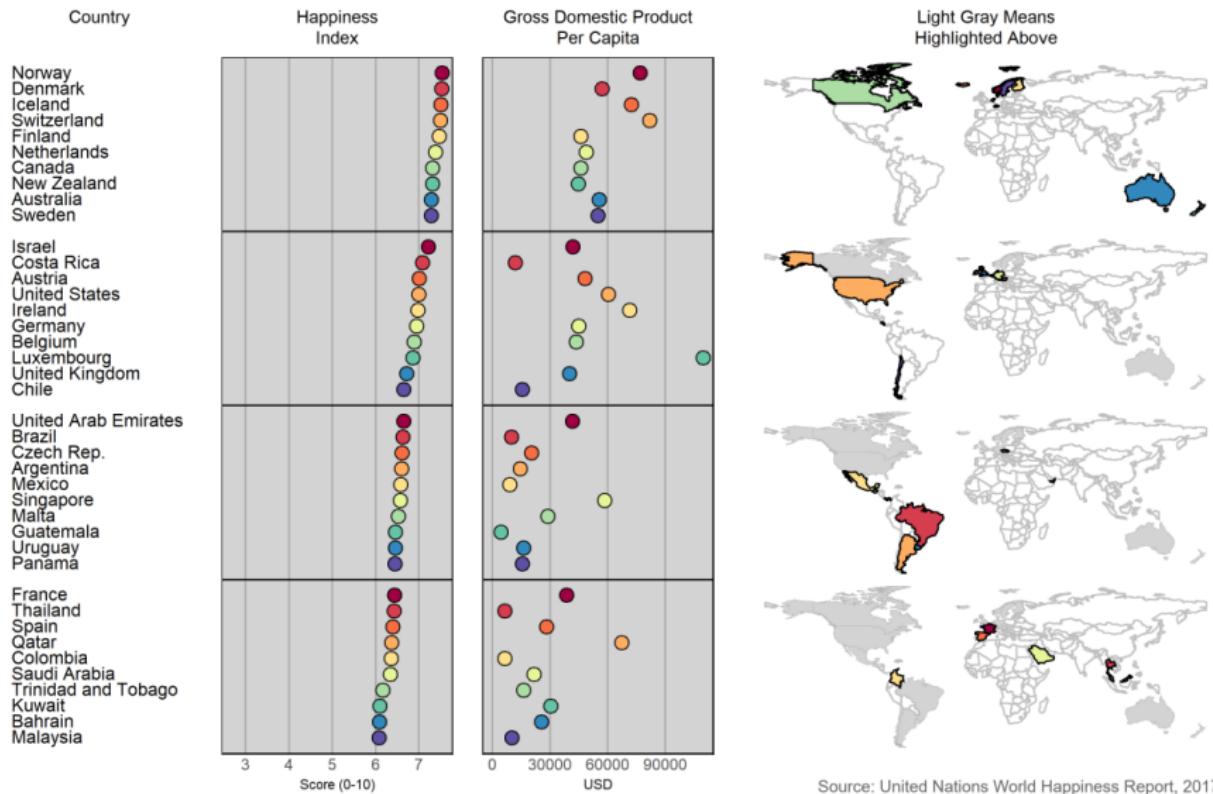


Image source (Few, 2014)

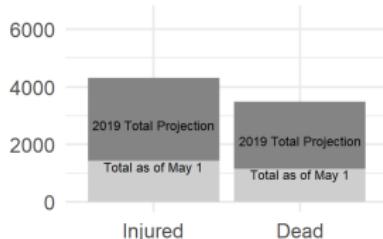
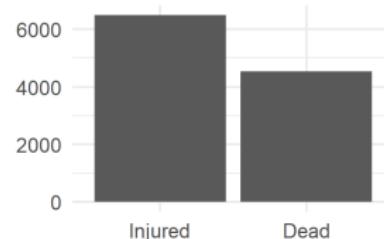
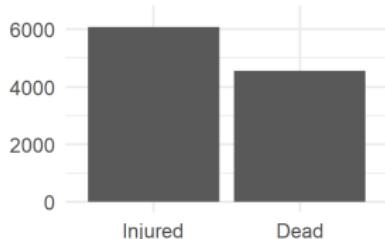
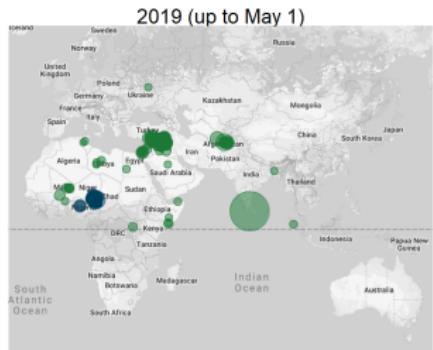
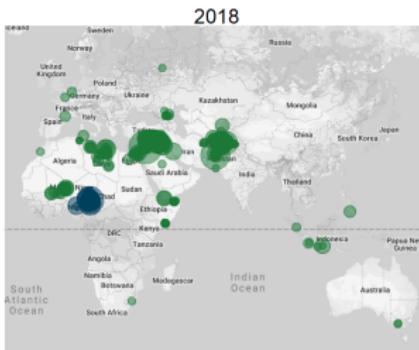
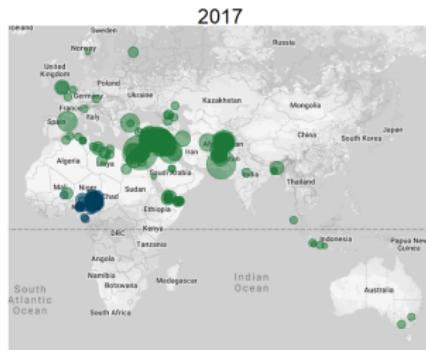
# Linked micromaps



# Proportional symbol map

Attacks claimed by the Islamic State and Boko Haram, 2017-2019

Point size is proportional to the number of deaths and injuries in each attack.  
ISIS in green, Boko Haram in blue.

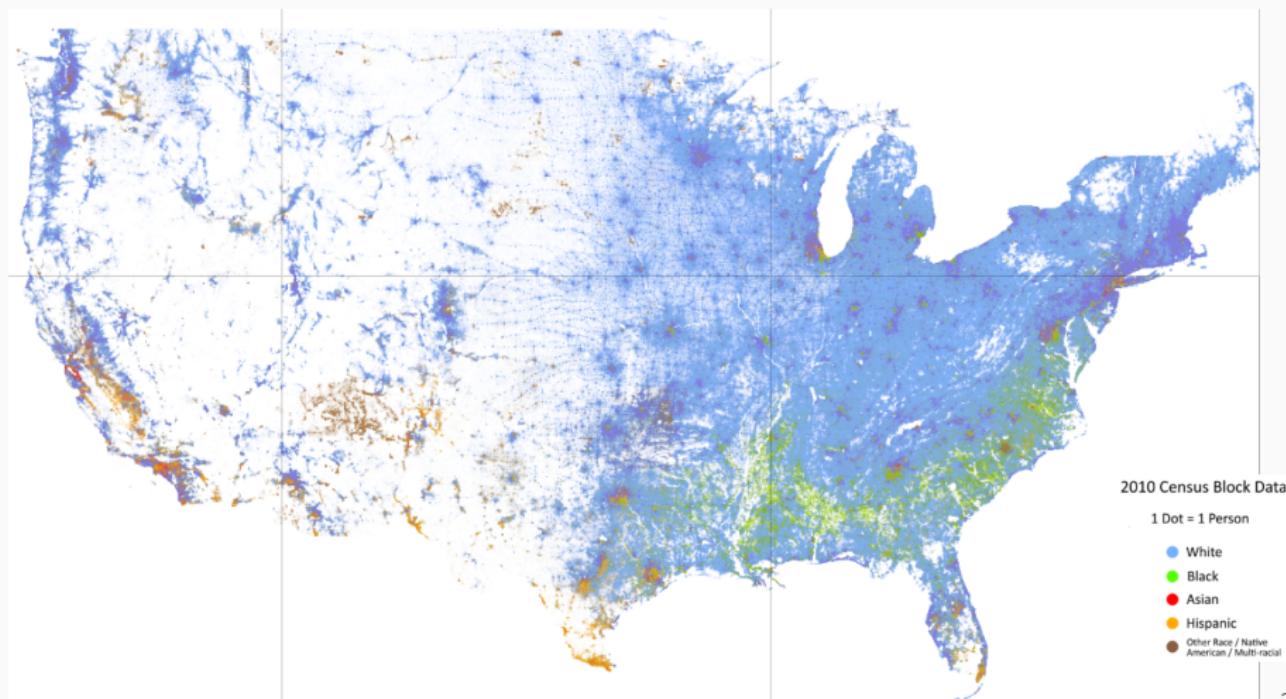


Source: Wikipedia, List of terrorist incidents

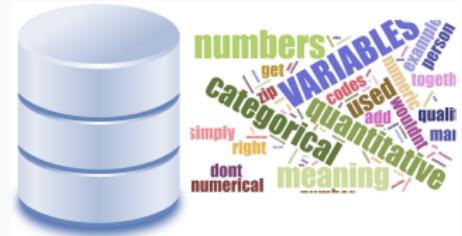
# Dot density map

Quantitative: Latitude, longitude

Categorical: Race/ethnicity

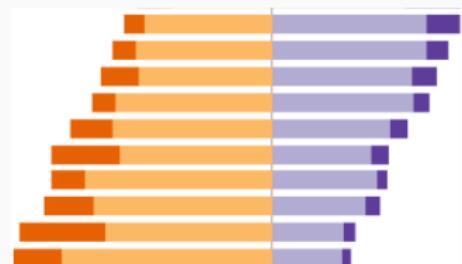


In summary, three precepts guide one's choice of graph type



numbers  
get  
zip  
simply  
don't  
numerical  
**VARIABLES**  
categorical  
quantitative  
meaning  
codes  
used  
add  
qualit  
mai  
exampl  
erson  
togeth  
right  
simply  
don't  
numerical

**Data structure** determines  
which graph types are feasible



**Graphical repertoire** determines  
which graph types you explore



**Design choices** align your visual argument  
with your rhetorical goals

# References

Few S (2014) Are mosaic plots worthwhile? *Visual Business Intelligence Newsletter*  
<https://tinyurl.com/y23majn2>

Harrelson N (2019) *Portfolio of data displays*. Rose-Hulman Institute of Technology, ME447 coursework

Hurley C (2019) *gclus: Clustering Graphics*. R package version 1.3.2  
<https://CRAN.R-project.org/package=gclus>

MIDFIELD (2019) Multiple-Institution Database for Investigating Engineering Longitudinal Development  
<https://engineering.purdue.edu/MIDFIELD>

R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria <https://www.R-project.org>

Ryan T (2019) *Portfolio of data displays*. Rose-Hulman Institute of Technology, ME447 coursework

Unwin A (2015) *GDAdata: Datasets for the book Graphical Data Analysis with R*. R package version 0.93  
<https://CRAN.R-project.org/package=GDAdata>