

Depicting Quantitative Data

Dimensionality:

data about running clubs

- **Univariate:** only one variable describes the data
 - number of members in each club
- **Bivariate:** two variables/attributes
 - numbers of male and female members in each club
- **Tri-variate:** three variables/attributes
 - number of men, women, average race finishing position for the club
- **Multivariate:** more than three variables/attributes
 - number of men, women, membership fees, colour, founding year, average race finishing position

The data

Club name: categorical

although note that an alphabetic ordering may be imposed, making the data ordinal

Number of members: quantitative

Number of women: quantitative

Number of men: quantitative

Membership fees: quantitative

Colour: categorical

Founding year: quantitative

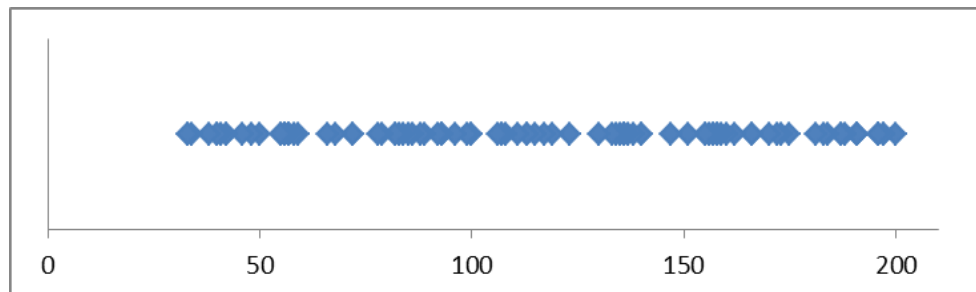
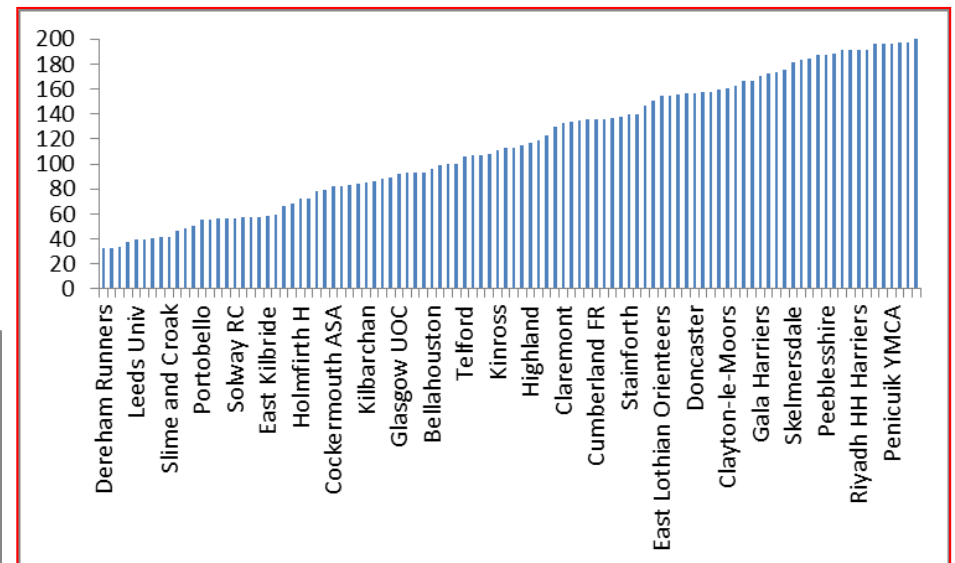
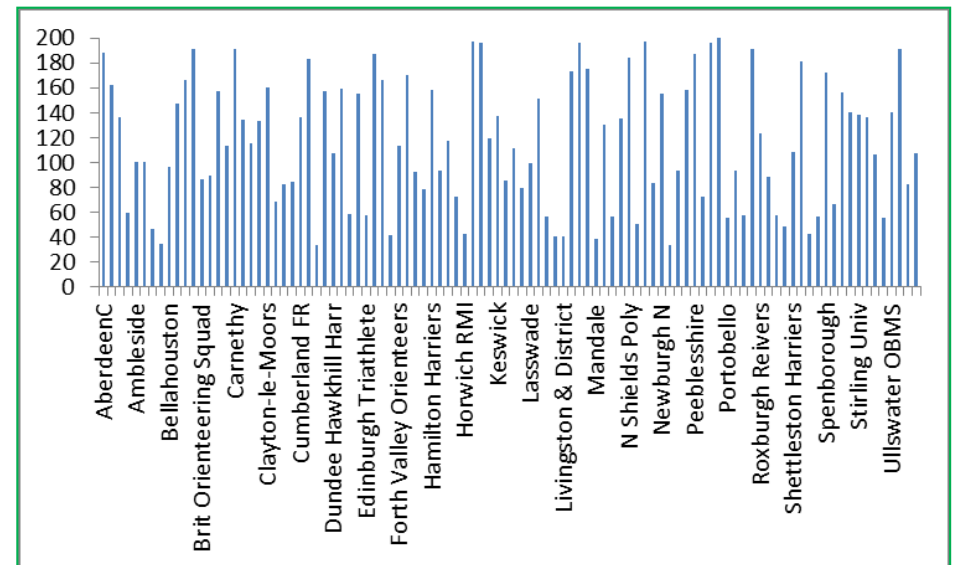
Average race finishing position: quantitative

Univariate

(number of members in a club)

AberdeenC	188
Achille Ratti	162
Aireborough Tri C	136
Alnwick H	59
Ambleside	100
Annan & Dist	100
Argyll & S	46
Ayr & Seaforth	34
Bellahouston	96
Black Isle	147
Border Harriers	166
Boundary H	191
Brit Orienteering Squad	86
Calderglen Harriers	89
Camuslang Harriers	157
Carnegie Harriers	113
Carnethy	191
Castlemilk	134
Central Region	115
Claremont	133
.... 100 clubs	

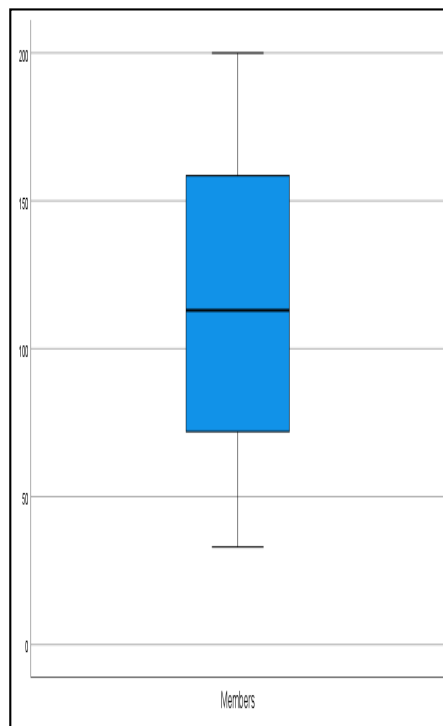
Perth Orienteers	200
Hunters Bog Trotters	197
New Town Hash	197
Inverness H	196
Lothian & Borders	196
Penicuik YMCA	196
Boundary H	191
Carnethy	191
Riyadh HH Harriers	191
Univeristy of Sunder	191
AberdeenC	188
Edinburgh Univ	187
Peeblesshire	187
N Shields Poly	184
Deeside Runners	183
Skelmersdale	181
Macclesfield H	175
Lochaber	173
Spenborough	172
Gala Harriers	170
.... 100 clubs	



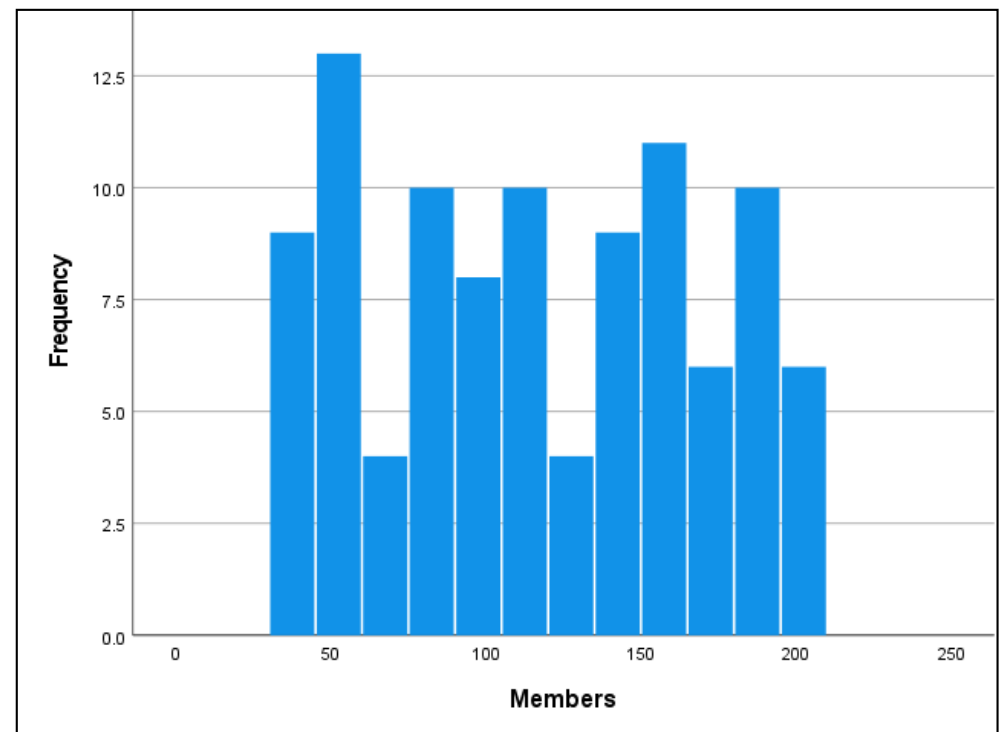
Univariate

(number of members in a club)

mean	116.2
std-dev	51.18
median	113
Q1	72
Q3	158.25

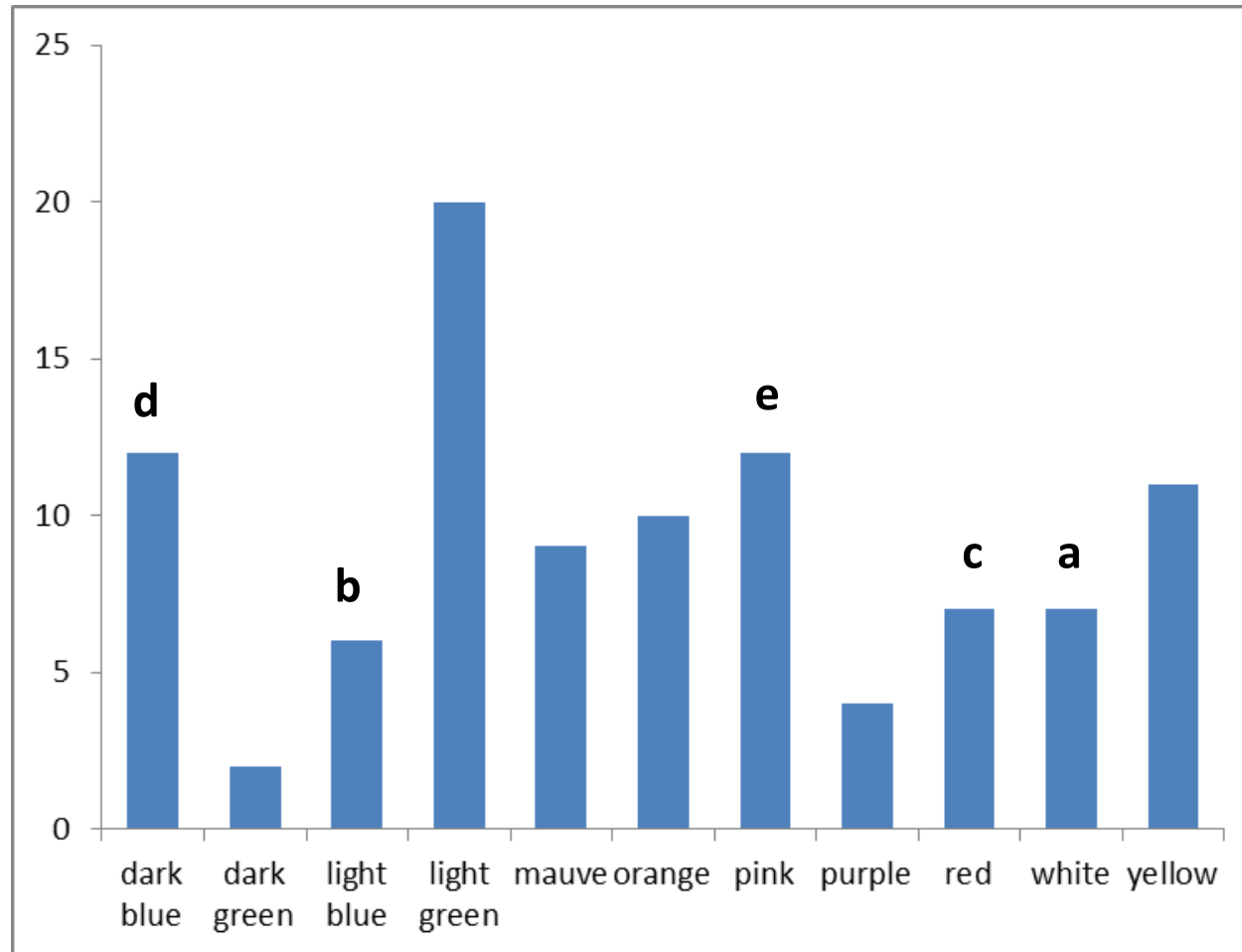


box plot



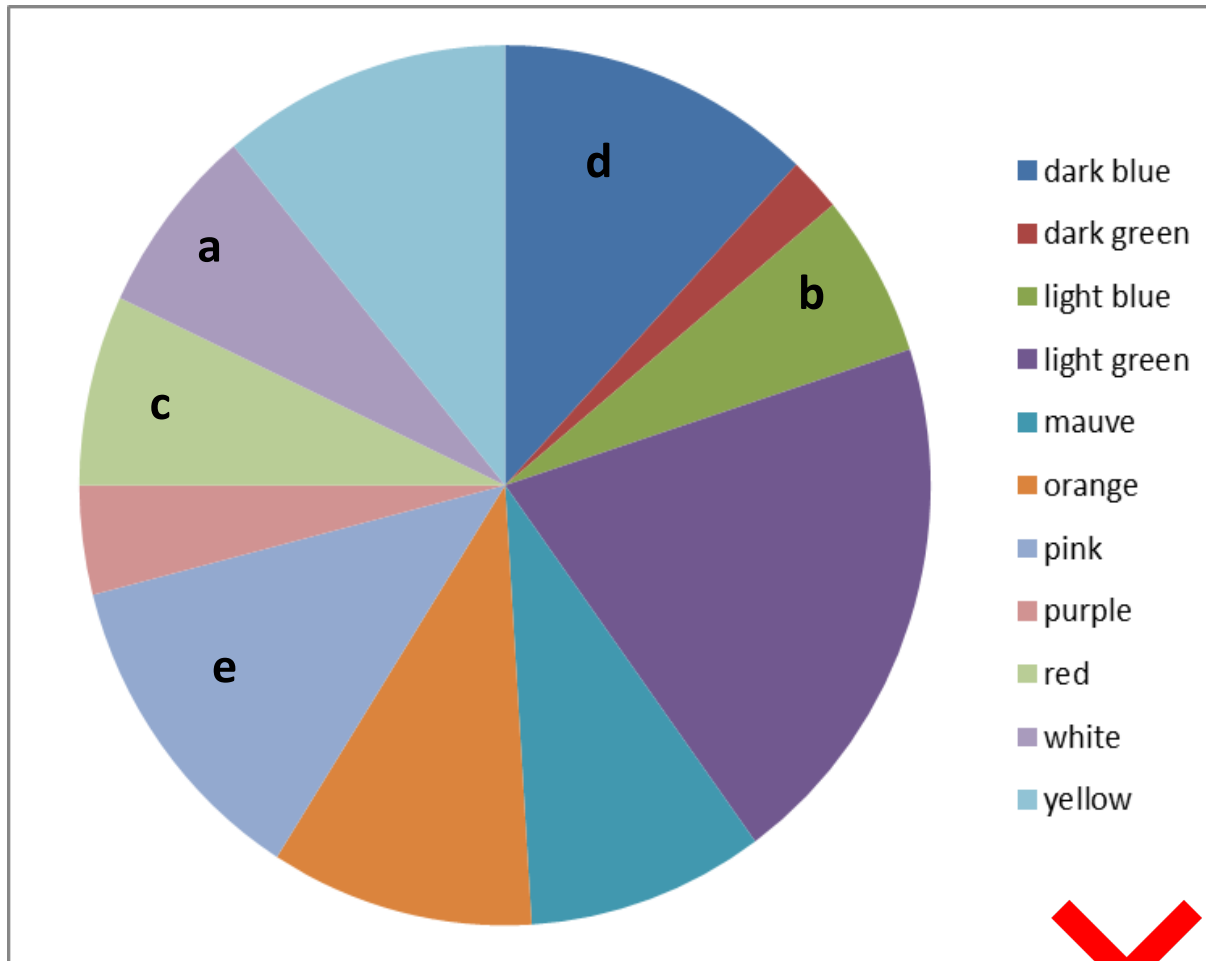
histogram

Bar charts



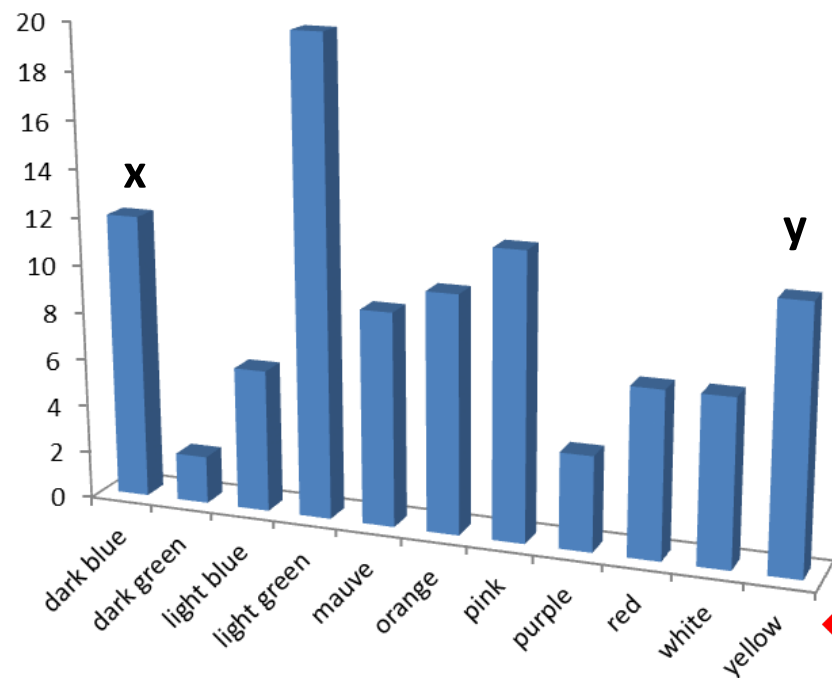
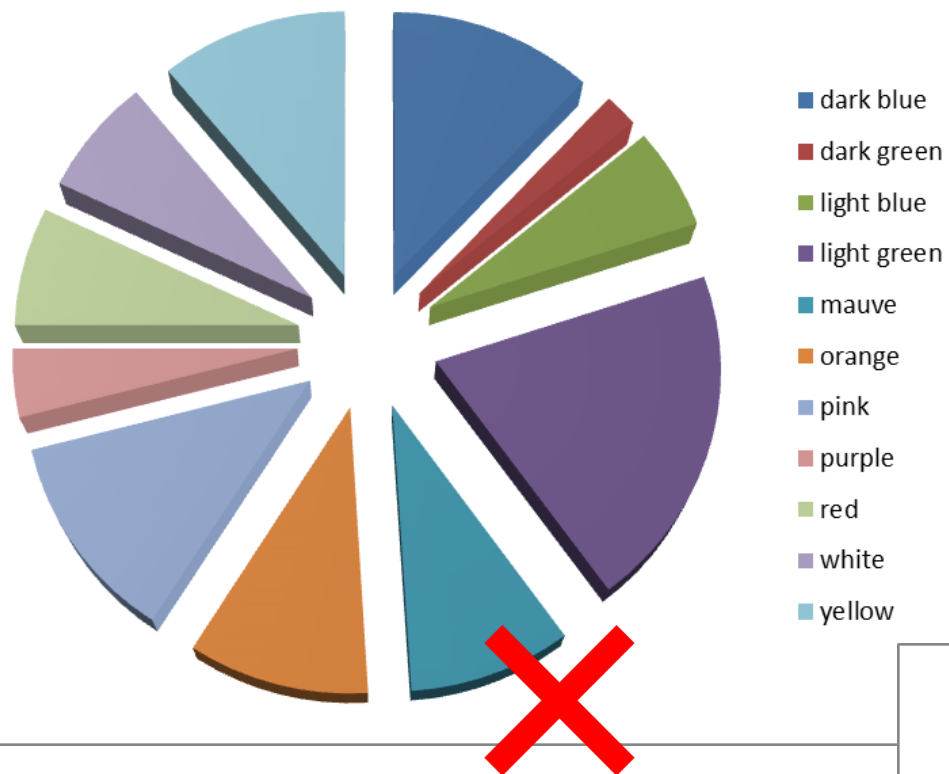
number of clubs associated with each colour

Pie charts are not good



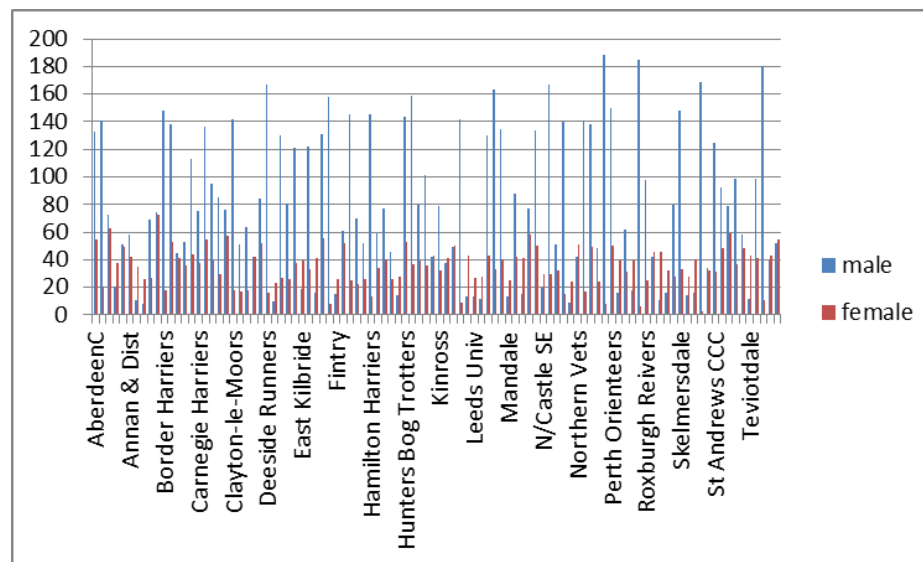
number of clubs associated with each colour

3D effects:
don't use
them!



AberdeenC	133	55
Achille Ratti	141	21
Aireborough Tri C	73	63
Alnwick H	21	38
Ambleside	51	49
Annan & Dist	58	42
Argyll & S	11	35
Ayr & Seaforth	8	26
Bellahouston	69	27
Black Isle	74	73
Border Harriers	148	18
Boundary H	138	53
Brit Orienteering Squad	45	41
Calderglen Harriers	53	36
Camuslang Harriers	113	44
Carnegie Harriers	75	38
Carnethy	136	55
Castlemilk	95	39
Central Region	85	30
Claremont	76	57
... 100 clubs		

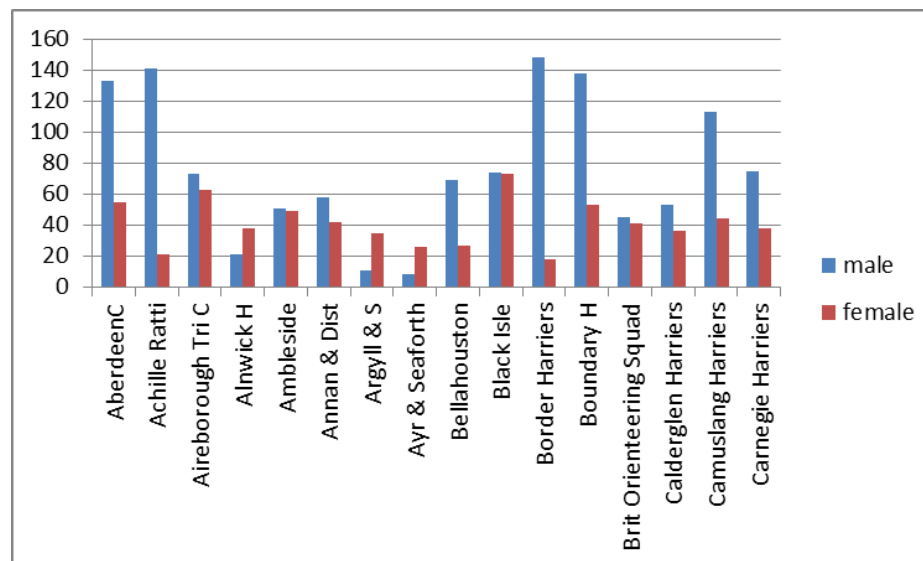
Bivariate (male and female members)

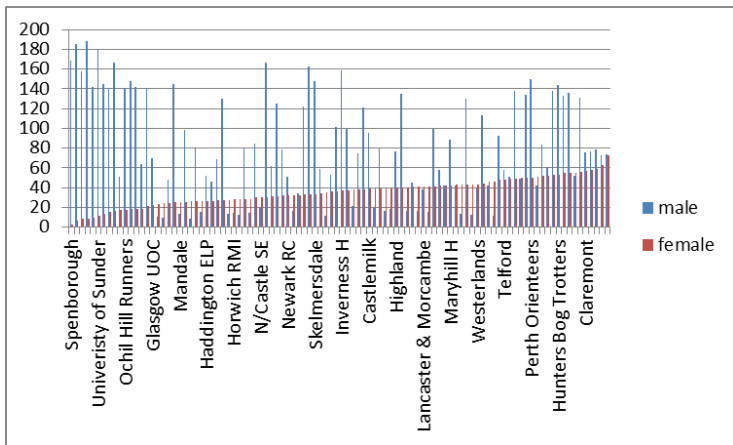


clustered bar chart (alphabetic)

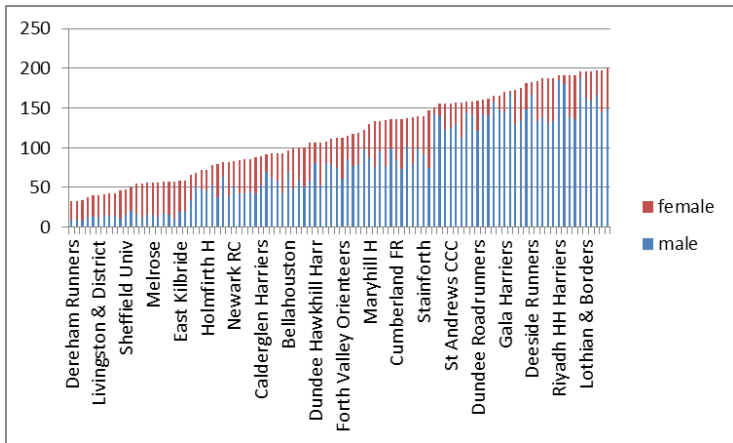
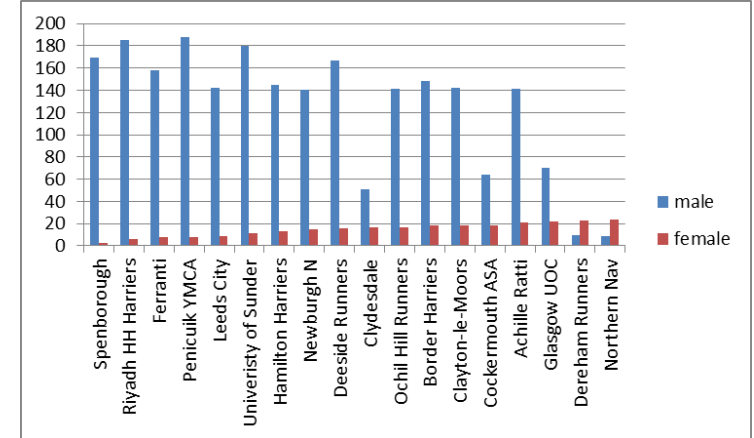
Overview of all clubs (top)

Detail of some clubs (bottom)

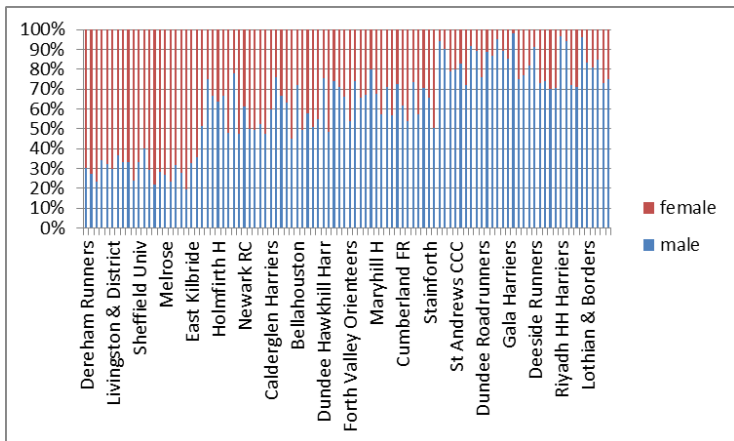
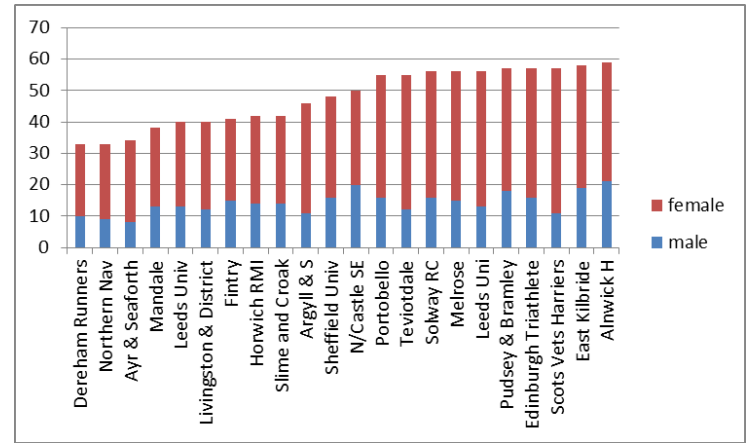




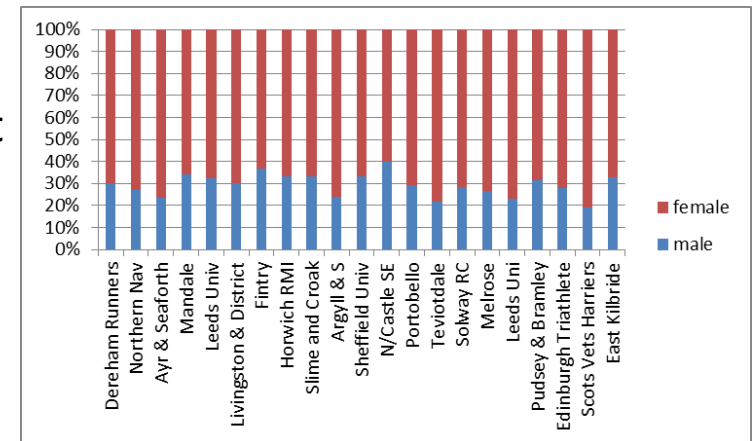
clustered bar chart
(ordered by female)

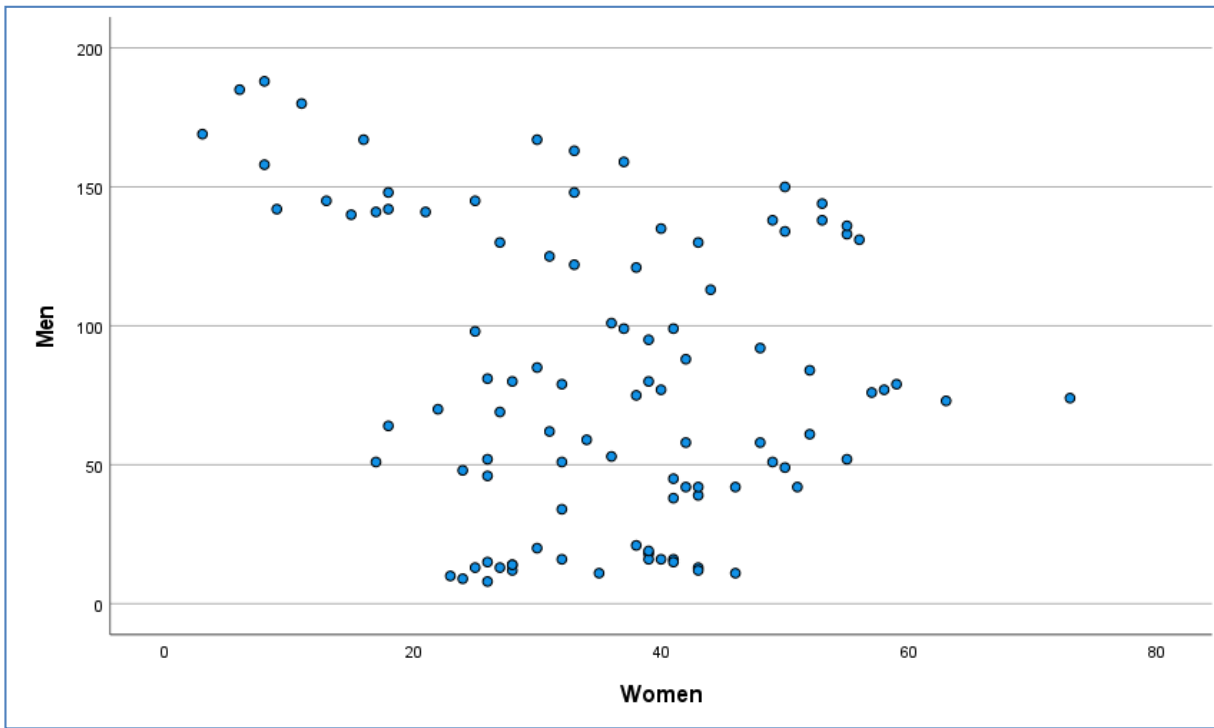


stacked bar chart
(ordered by total)

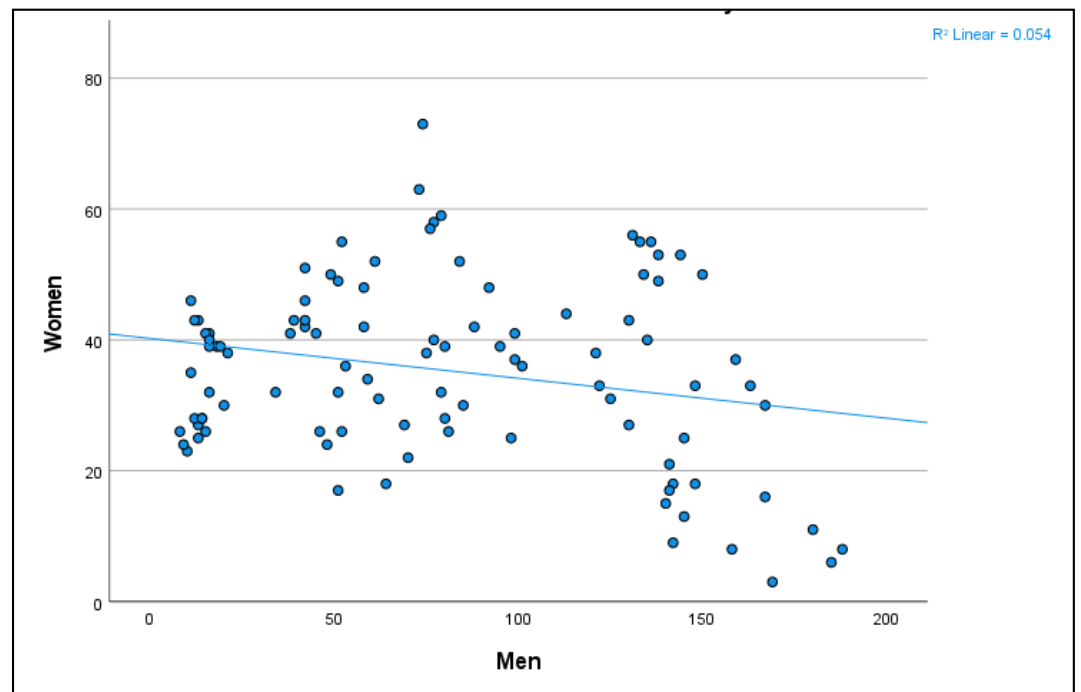


100% stacked bar chart
(ordered by total)

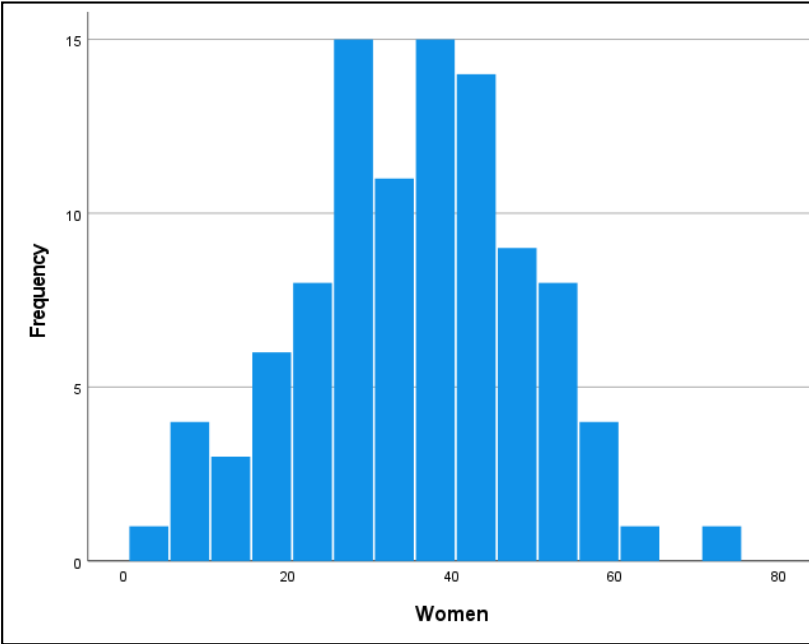
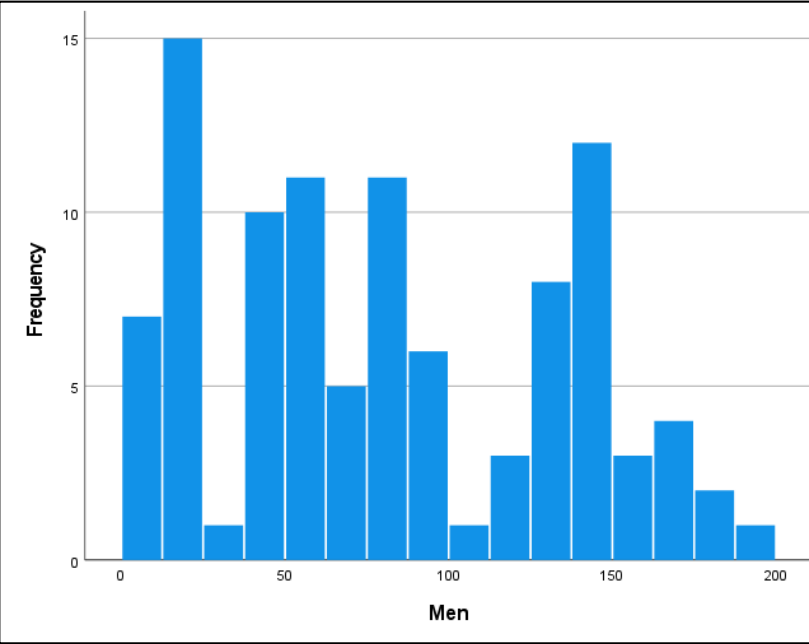
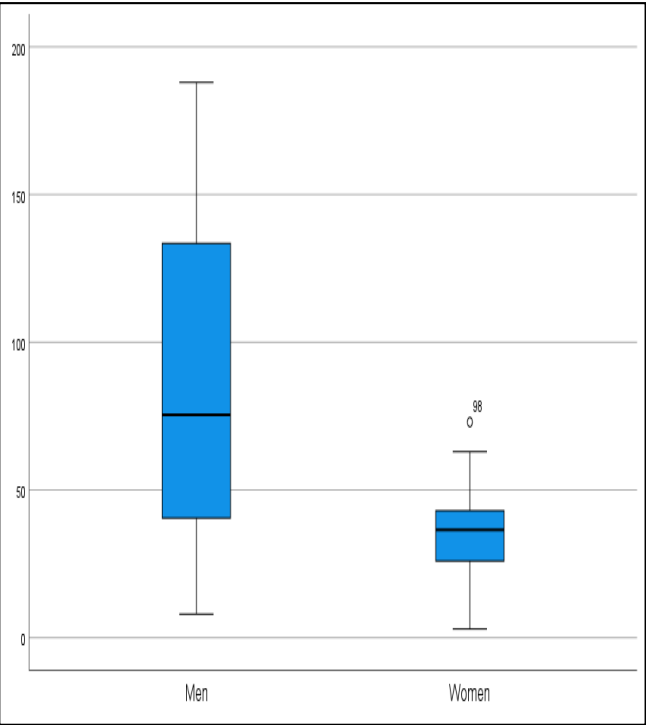




scatterplot

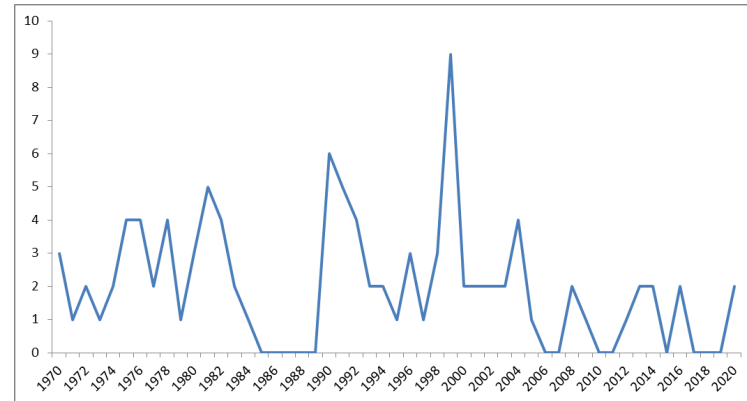
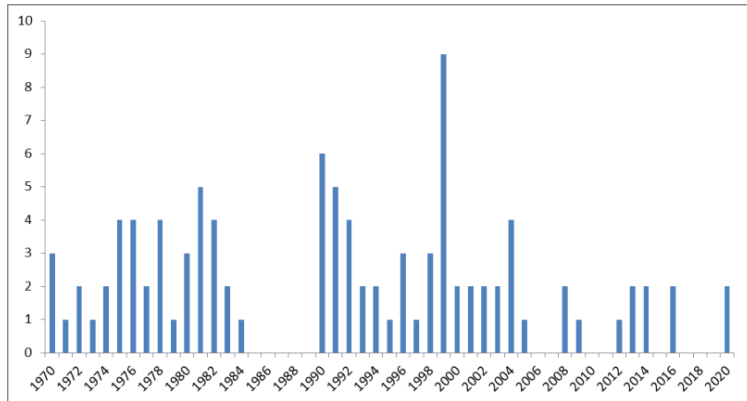


	male	female
mean	80.9	35.3
std-dev	52.6	13.8
median	75.5	36.5
Q1	41.3	26.0
Q3	133.3	43.0

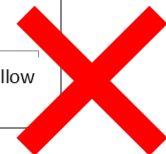
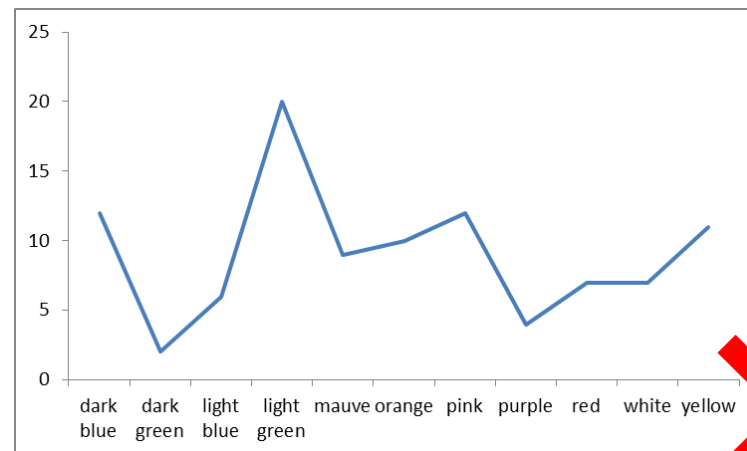
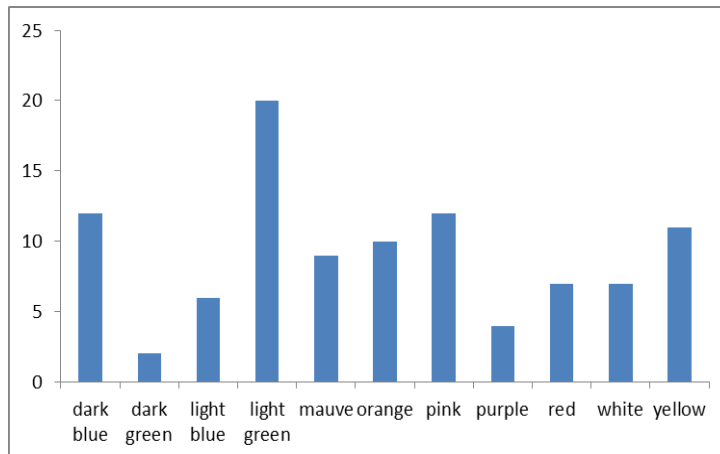


Bar charts vs Line charts

number of new clubs opened each year

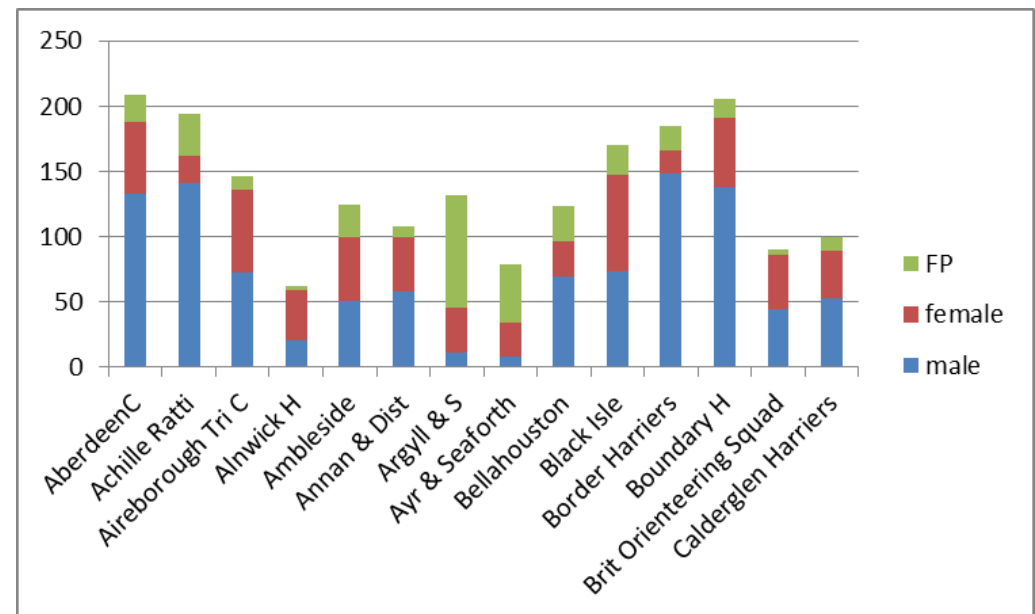
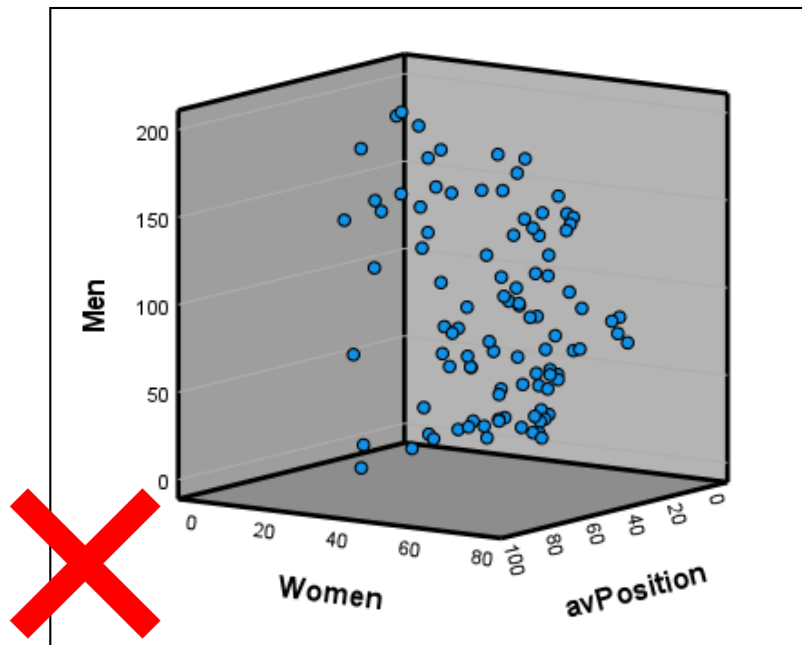
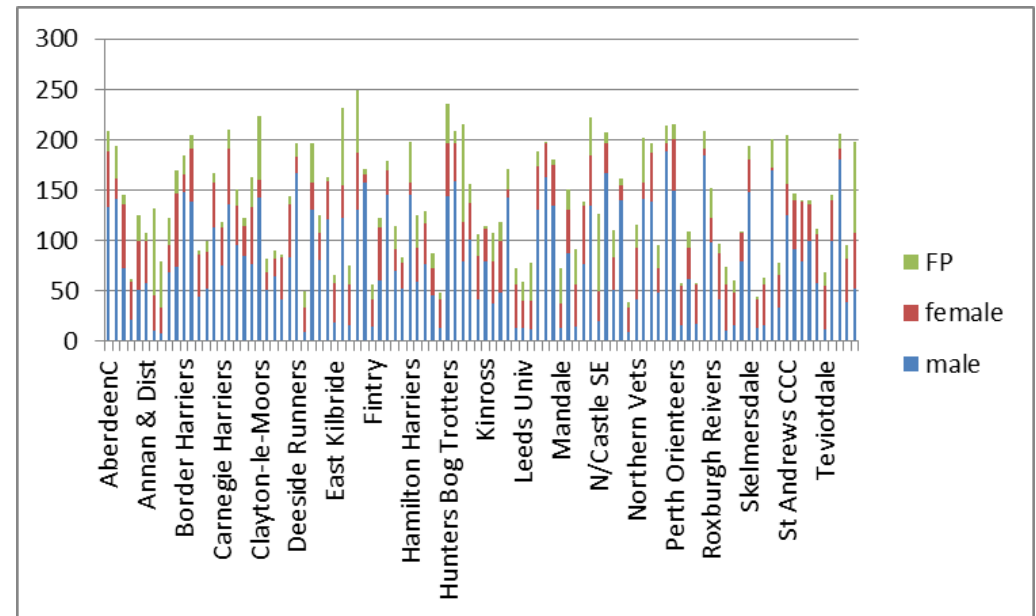


number of clubs associated with each colour

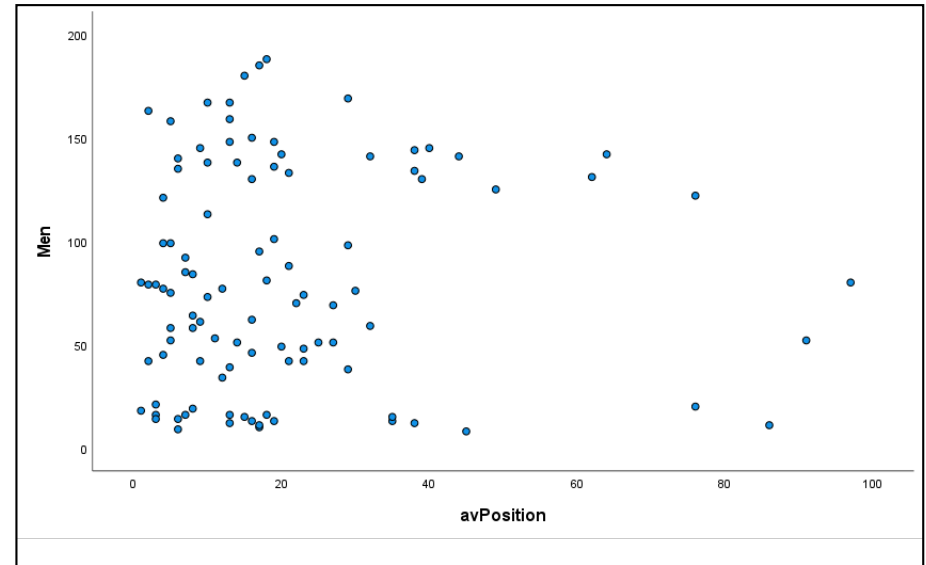
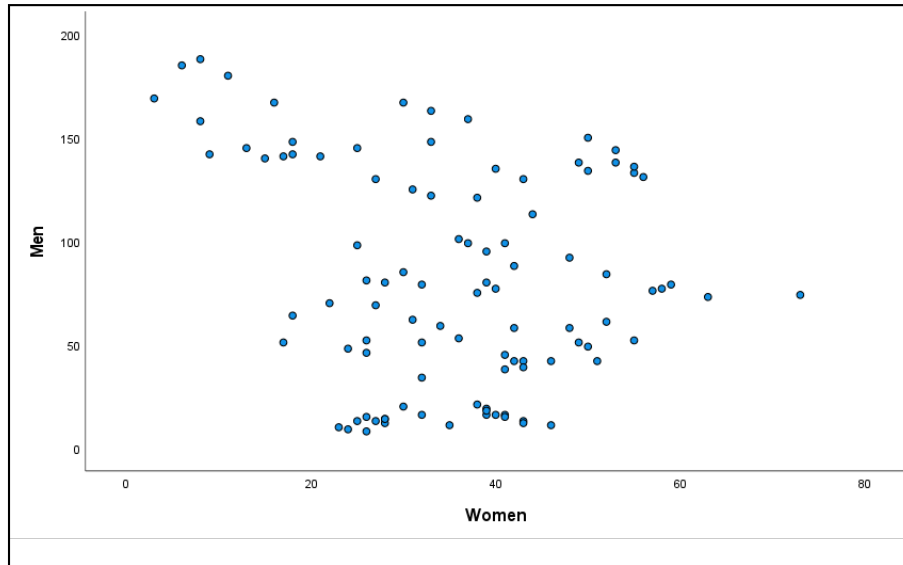


Tri-variate

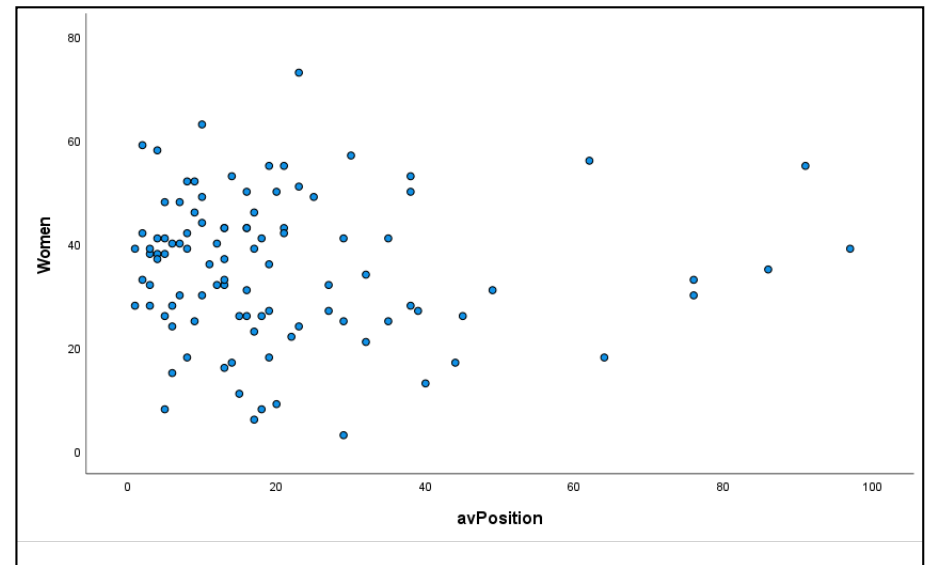
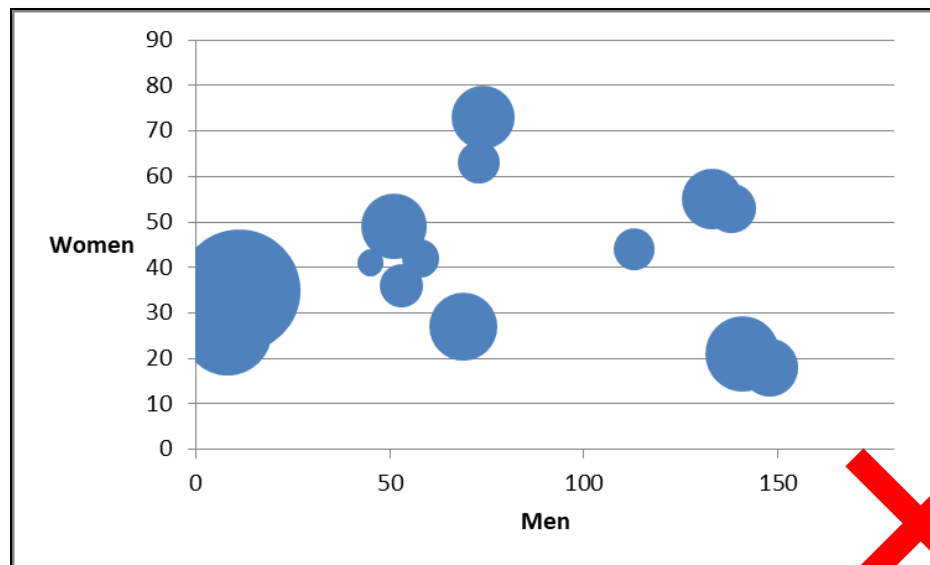
	male	female	average finishing position
AberdeenC	133	55	21
Achille Ratti	141	21	32
Aireborough Tri C	73	63	10
Alnwick H	21	38	3
Ambleside	51	49	25
Annan & Dist	58	42	8
Argyll & S	11	35	86
Ayr & Seaforth	8	26	45
Bellahouston	69	27	27
Black Isle	74	73	23
Border Harriers	148	18	19
Boundary H	138	53	14
Brit Orienteering Squad	45	41	4
Calderglen Harriers	53	36	11
Camuslang Harriers	113	44	10
... 100 clubs			



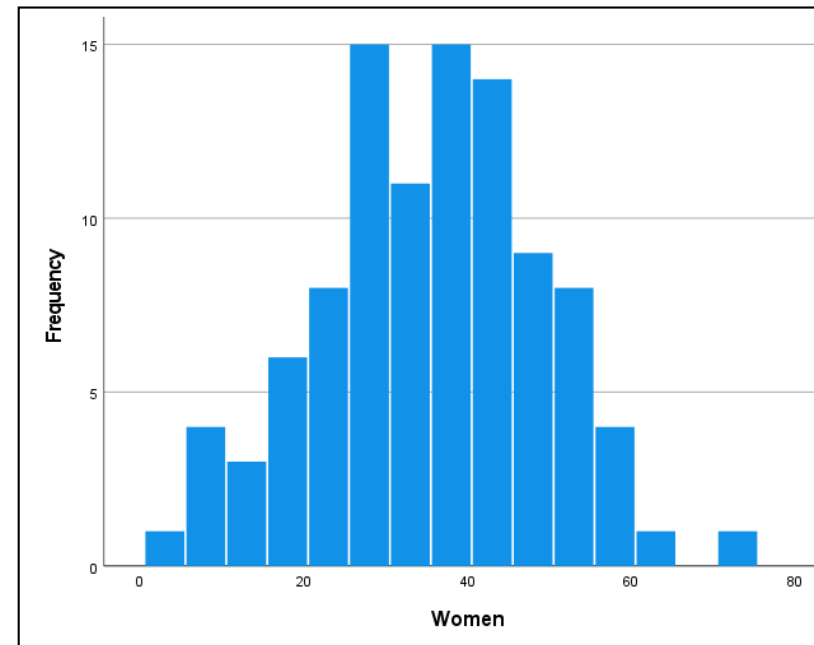
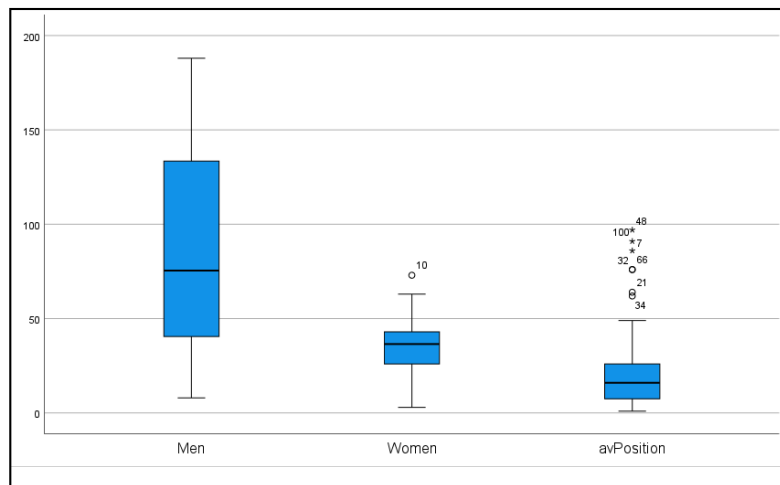
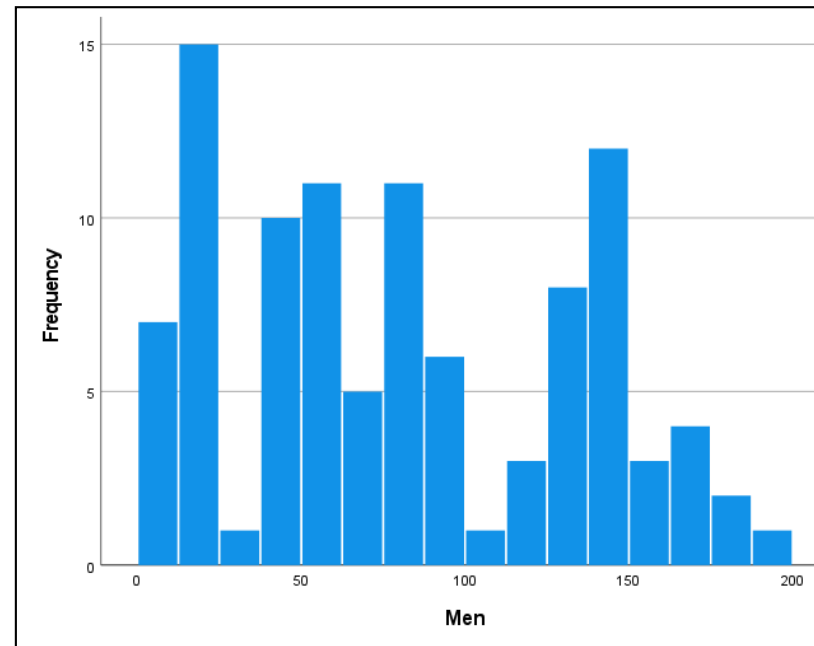
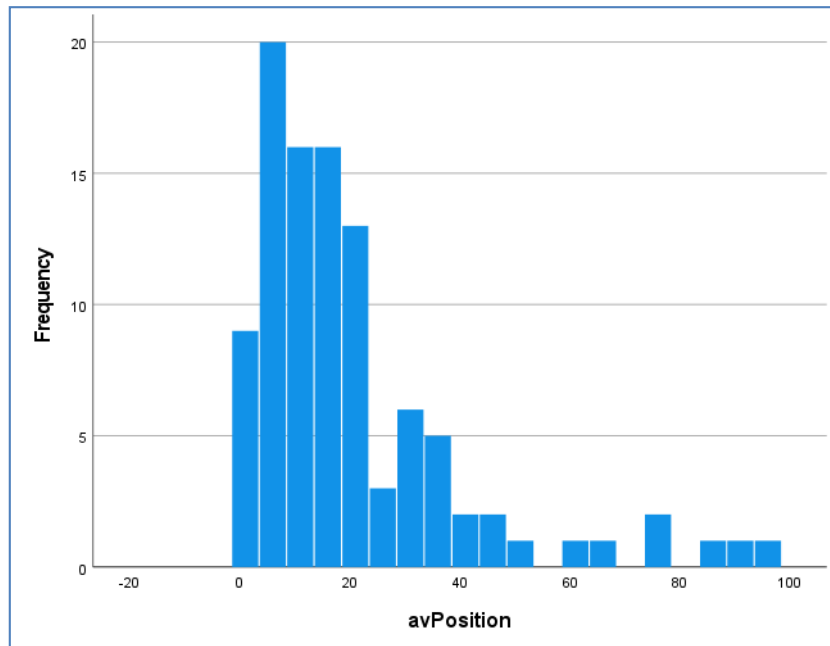
scatterplot matrix (SPLOM): one scatterplot for each pair of variables



bubble plot



One bar chart per variable: relationships between them best shown via interaction methods



Tri-variate: Heat maps

- Typically, two (independent) **categorical** variables, and a **quantitative** variable
- The categories are on the two axes
- The quantitative value is represented by change in colour value
 - typically: ‘darker’ = ‘more’... but be careful!
 - Often best to show the value encoding in a chart legend
- The order of the categories on each axis can be changed (and may be important for identification of patterns)
- Each cell has only one value

	trivial	easy	medium	hard	guelling
Jan					
Feb					
Mar					
Apr					
May					
Jun					
Jul					
Aug					
Sep					
Oct					
Nov					
Dec					

Record finishing time for races over the same distance,
with different difficulty, at different times of year

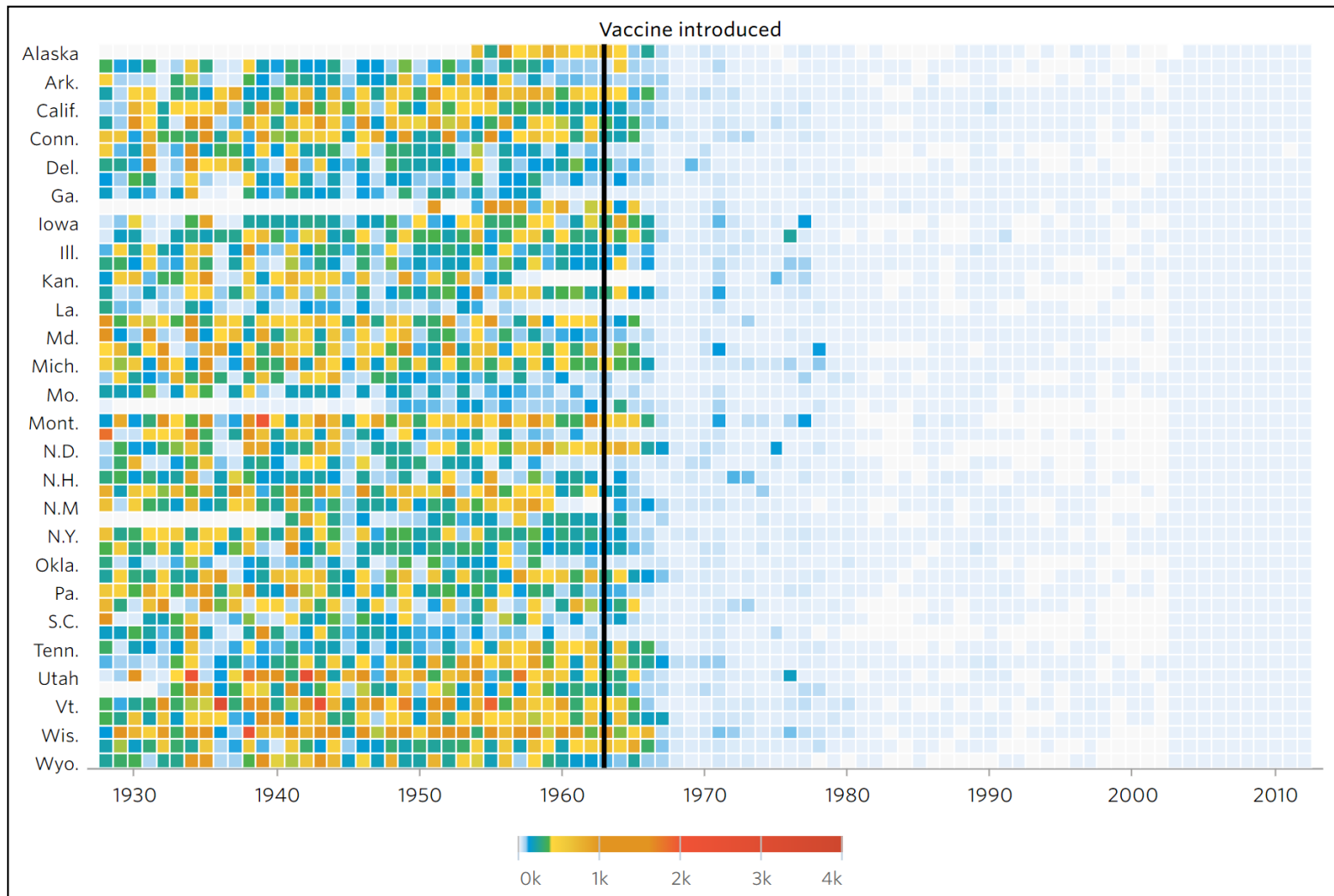
	Jenny	Hannah	Chen	Thembi	Farah
A					
B					
C					
D					
E					
F					
G					
H					
I					
J					
K					
L					

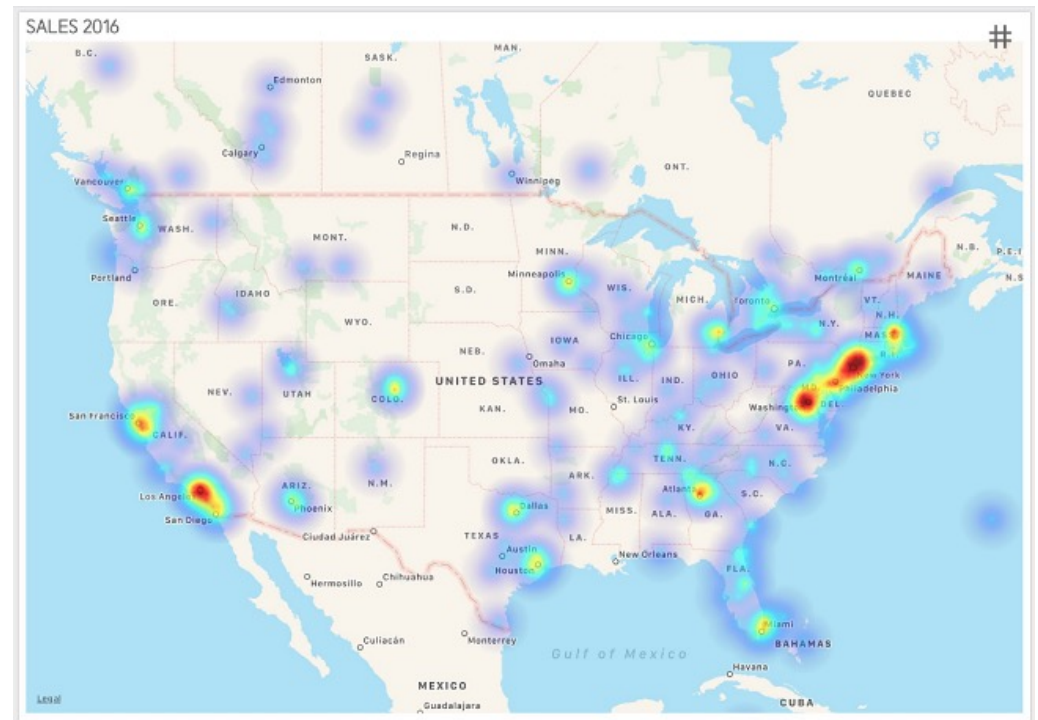
Proportion of baby girls given particular names,
with respect to different countries

	Jenny	Hannah	Chen	Thembi	Farah
F					
L					
G					
C					
B					
A					
J					
E					
I					
H					
D					
K					

Proportion of baby girls given particular names,
with respect to different countries, reordered

Measles cases over time, per US state



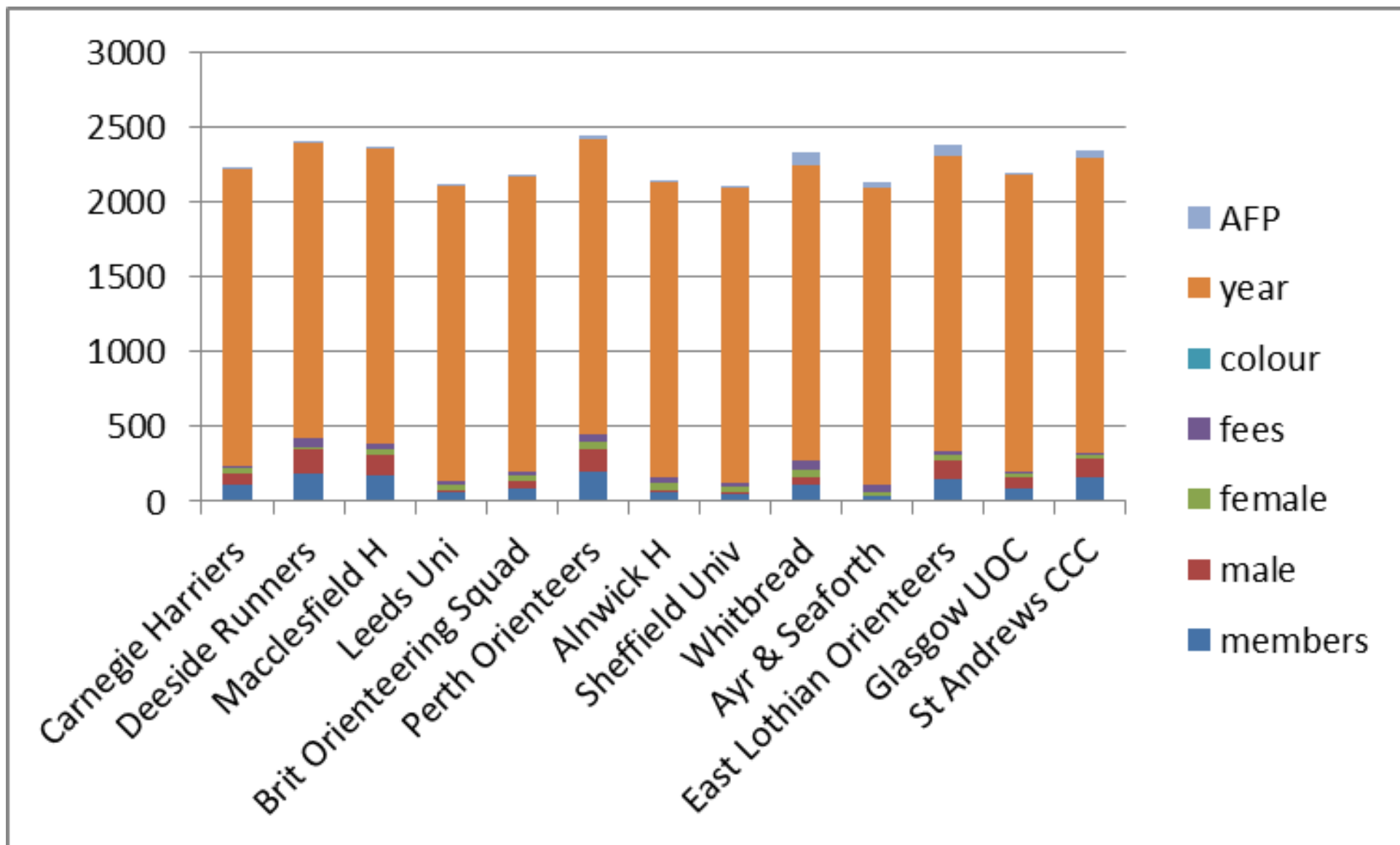


<https://onlygrowth.com/blogs/posts/how-to-use-heat-maps-and-eye-tracking-software-to-improve-ux-and-lift-conversion-rates> (accessed 25/05/21)

<https://www.infragistics.com/community/blogs/b/mobileman/posts/geographical-heat-maps-and-how-to-use-them-with-reportplus> (accessed 25/05/21)

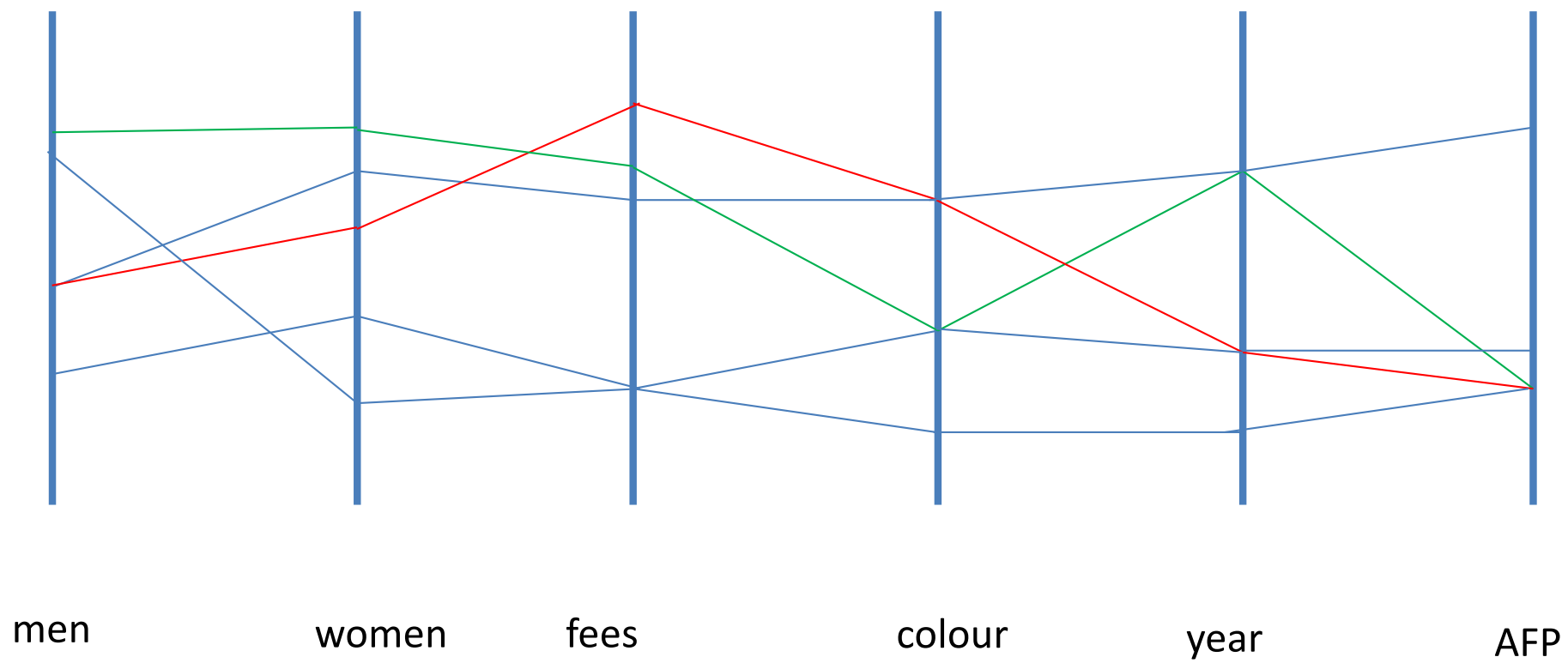
Multivariate

	members	male	female	fees	colour	year	average finishing position
Carnegie Harriers	113	75	38	15	dark blue	1970	5
Deeside Runners	183	167	16	60	yellow	1970	13
Macclesfield H	175	135	40	40	light green	1970	6
Leeds Uni	56	13	43	20	dark green	1971	16
Brit Orienteering Squad	86	45	41	25	light green	1972	4
Perth Orienteers	200	150	50	50	light green	1972	16
Alnwick H	59	21	38	40	light green	1973	3
Sheffield University	48	16	32	25	white	1974	13
Whitbread	107	52	55	55	yellow	1974	91
Ayr & Seaforth	34	8	26	45	yellow	1975	45
East Lothian Orienteers	155	122	33	20	yellow	1975	76
Glasgow UOC	92	70	22	15	purple	1975	22
St Andrews CCC	156	125	31	10	yellow	1975	49
Doncaster	157	130	27	50	red	1976	39
Edinburgh Triathlete	57	16	41	40	dark blue	1976	18
Forth Valley Orienteers	113	61	52	15	purple	1976	9
Scots Vets Harriers	57	11	46	30	light green	1976	17
Keswick	137	101	36	40	light green	1977	19
Lasswade	99	49	50	35	light green	1977	20
Calderglen Harriers	89	53	36	30	pink	1978	11



Multivariate: Parallel coordinates

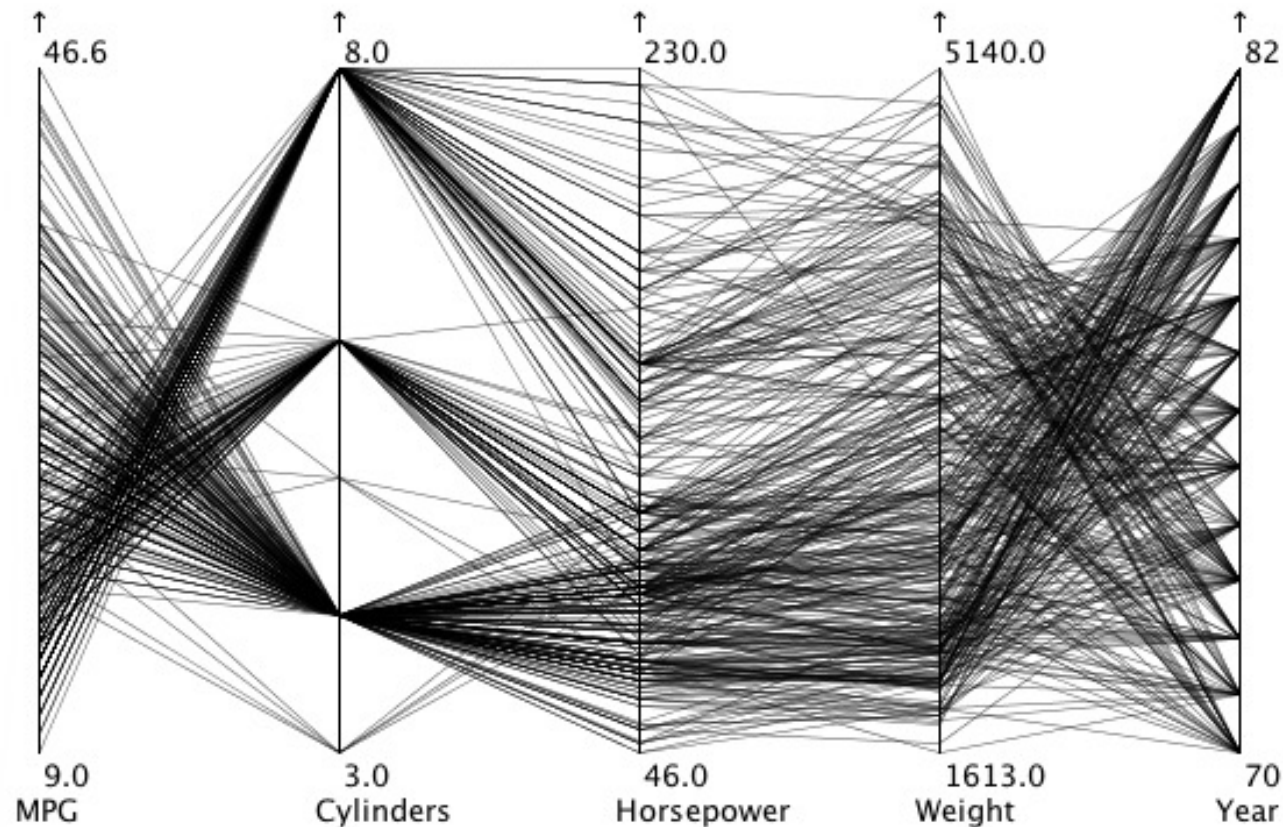
- Each vertical axis is a dimension, with its values equally spaced along it
- The dimensions are arranged, equally spaced, horizontally
- A single data point is a line that joins its values on each dimension



Car models:

- released from 1970 to 1982
- mileage (MPG)
- number of cylinders
- horsepower
- weight
- year
- *(plus other features not used here)*

MPG	Cylinders	Horsepower	Weight	Year
15	8	170	3563	1970
14	8	160	3609	1970
15	8	150	3761	1970
14	8	225	3086	1970
24	4	95	2372	1970
22	6	95	2833	1970
18	6	97	2774	1970
21	6	85	2587	1970
27	4	88	2130	1970
26	4	46	1835	1970
25	4	87	2672	1970
24	4	90	2430	1970
25	4	95	2375	1970
...and many more data items				



Each line from left to right represents one car

Looking at each pair of axes in turn:

- the cylinder axis has only a few values – all lines pass through a small number of points
- 8-cylinder cars tend to have lower mileage than cars with 6 or 4 cylinders (inverse correlation)
- more cylinders means more horsepower (almost direct correlation)
- more horsepower means more weight (almost direct correlation)
- older cars are heavier (roughly, an inverse correlation)

Parallel coordinate transformations

The iris data set

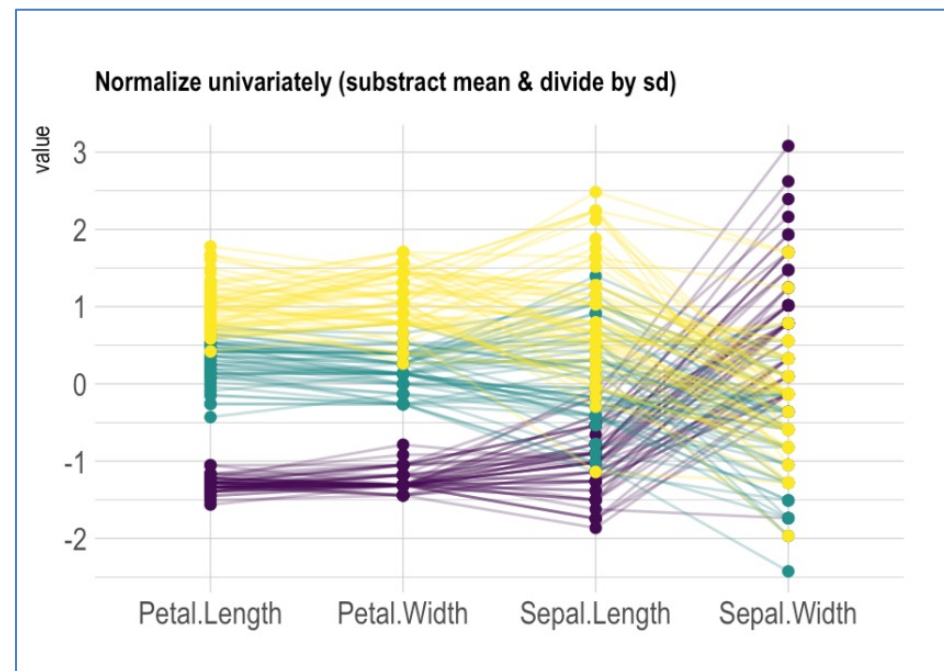
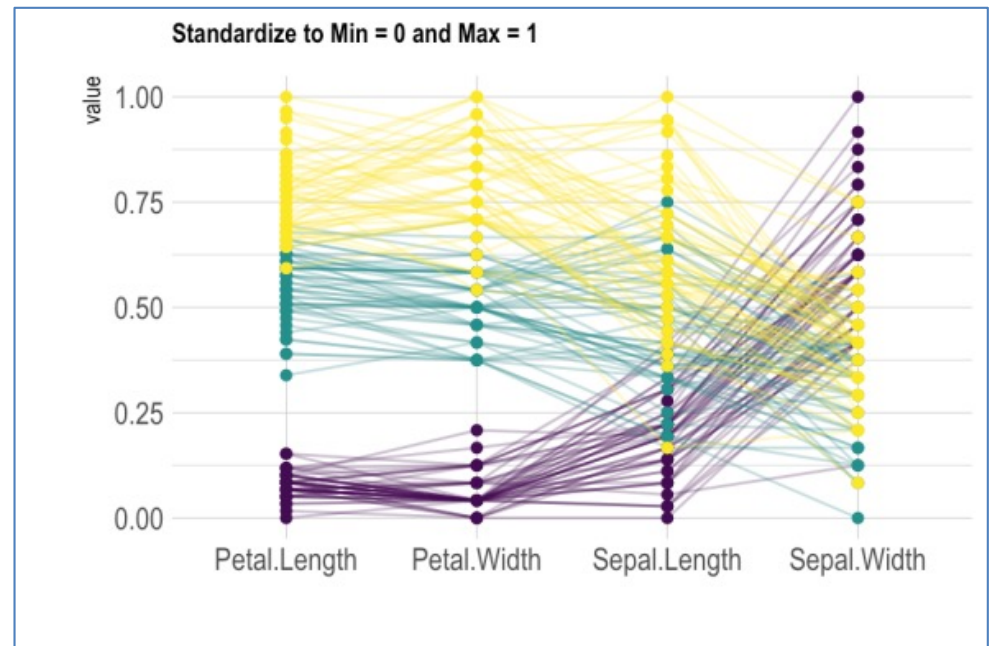
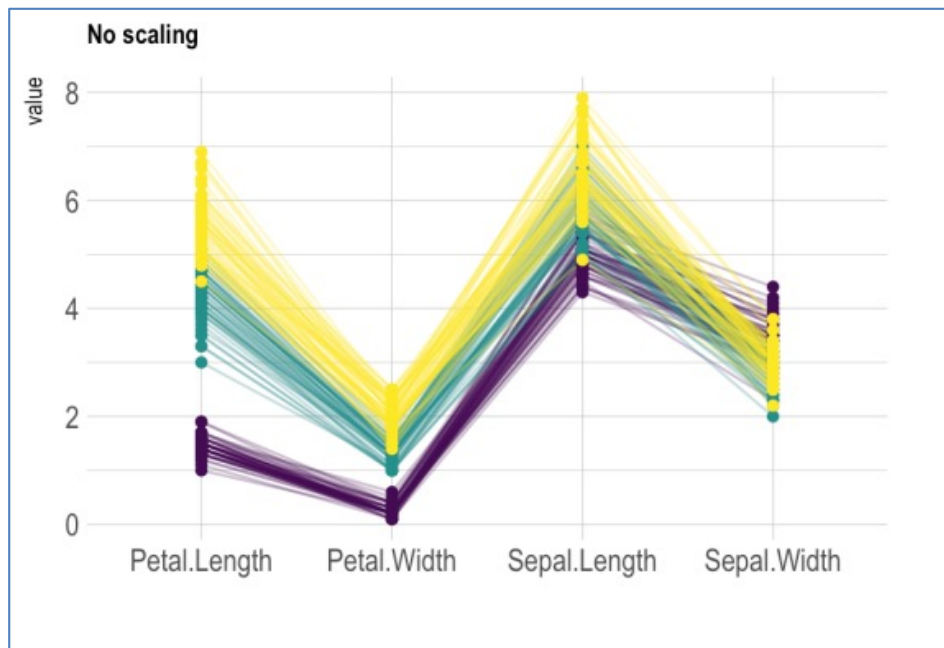
4 dimensions

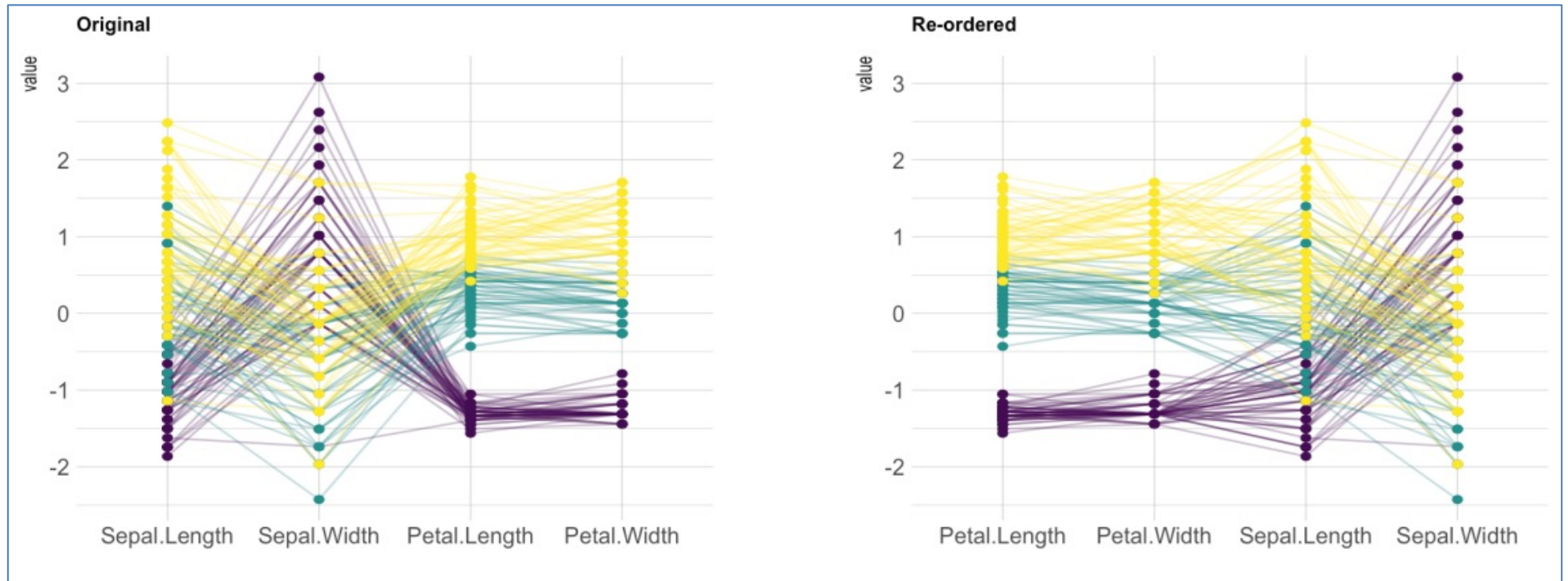
- petal width
- petal length
- sepal width
- sepal length



4.8, 3.0, 1.4, 0.3, Iris-setosa
5.1, 3.8, 1.6, 0.2, Iris-setosa
5.3, 3.7, 1.5, 0.2, Iris-setosa
5.0, 3.3, 1.4, 0.2, Iris-setosa
7.0, 3.2, 4.7, 1.4, Iris-versicolor
6.4, 3.2, 4.5, 1.5, Iris-versicolor
6.9, 3.1, 4.9, 1.5, Iris-versicolor
5.1, 2.5, 3.0, 1.1, Iris-versicolor
5.7, 2.8, 4.1, 1.3, Iris-versicolor
6.3, 3.3, 6.0, 2.5, Iris-virginica
5.8, 2.7, 5.1, 1.9, Iris-virginica
7.1, 3.0, 5.9, 2.1, Iris-virginica
6.3, 2.9, 5.6, 1.8, Iris-virginica

.....

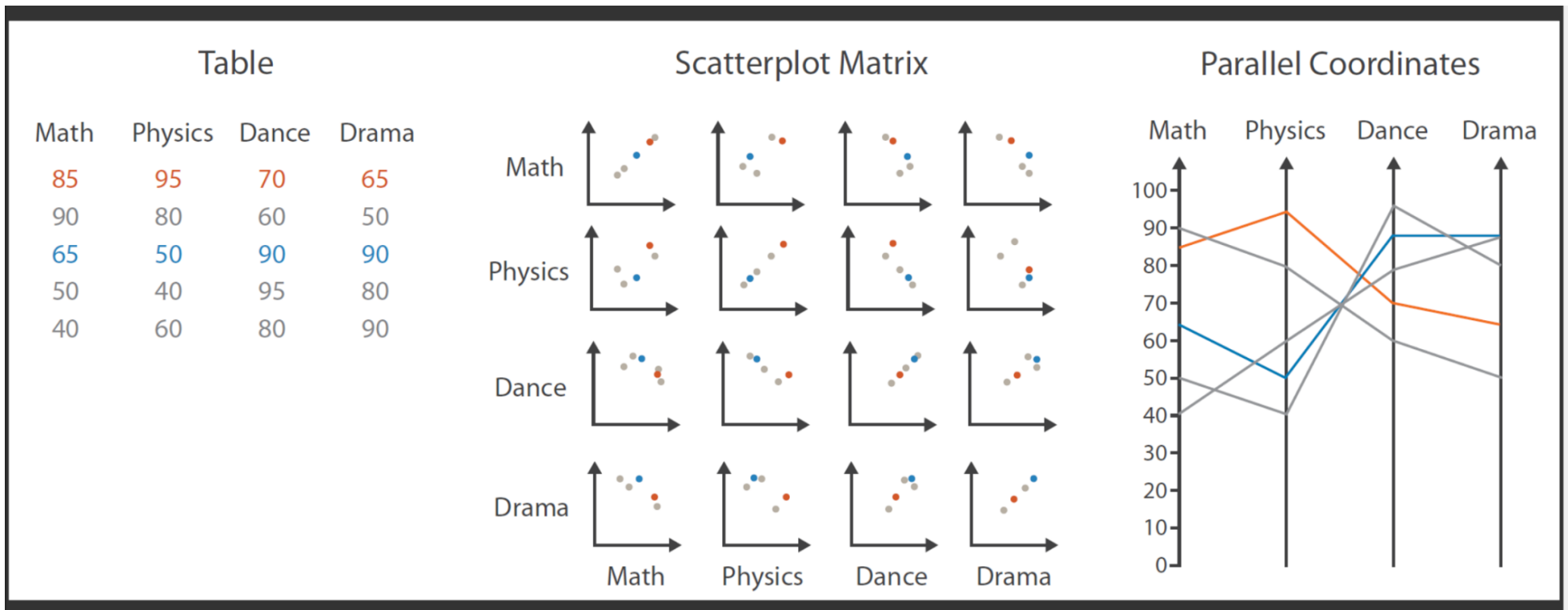




Arrange the order of the dimensions on the x-axis to find and/or highlight clear relationships (direct or inverse) of interest

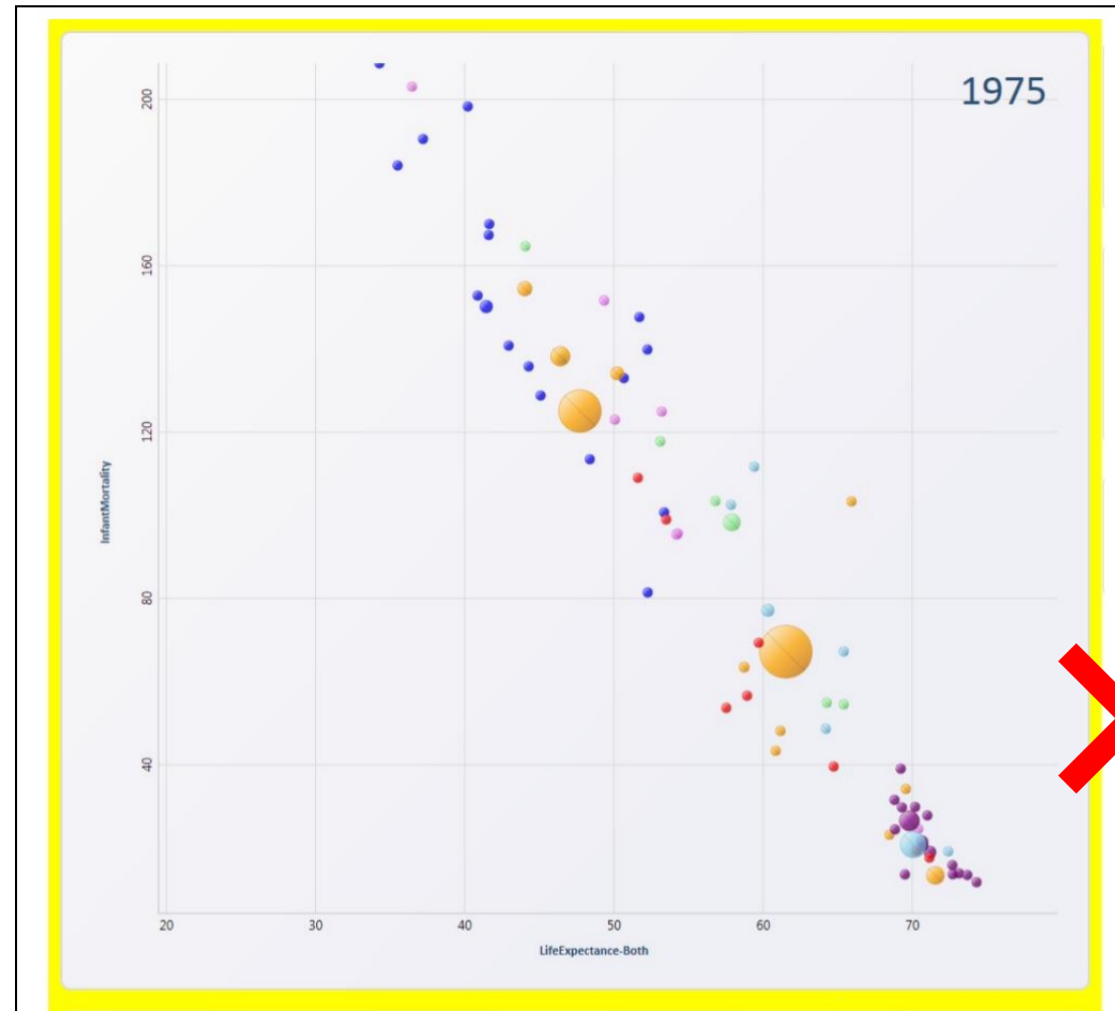
Multivariate:

Scatterplot matrix & Parallel coordinates

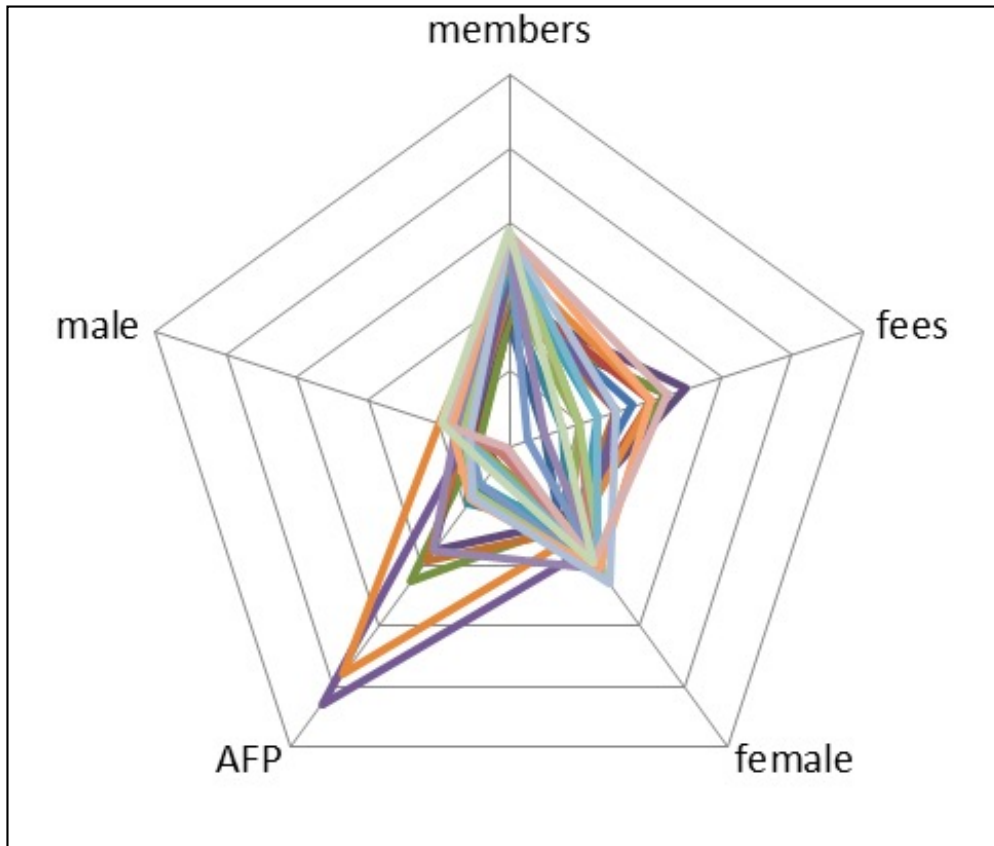


Multivariate: Bubble Plots

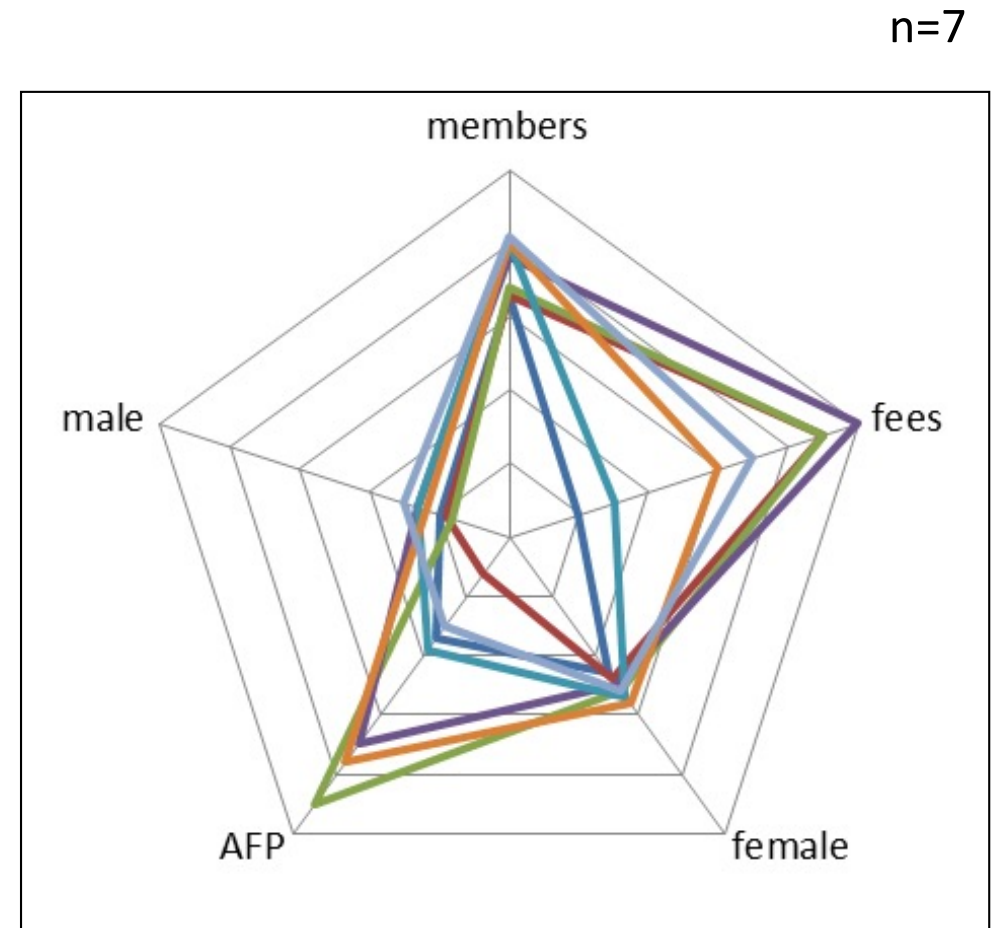
x axis: life expectancy
y axis: infant mortality
size: population
colour: continent



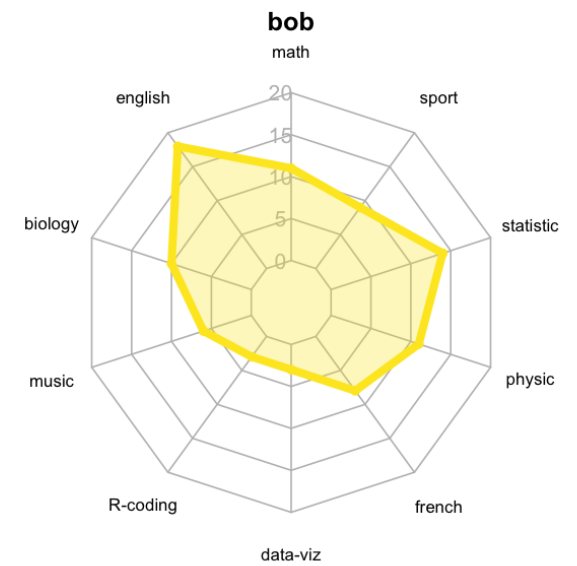
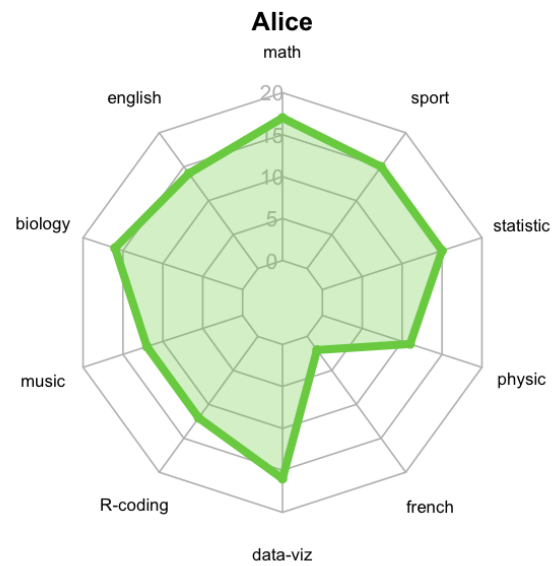
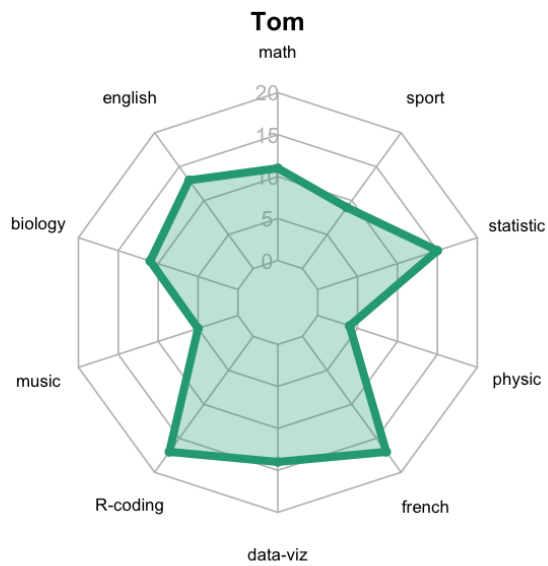
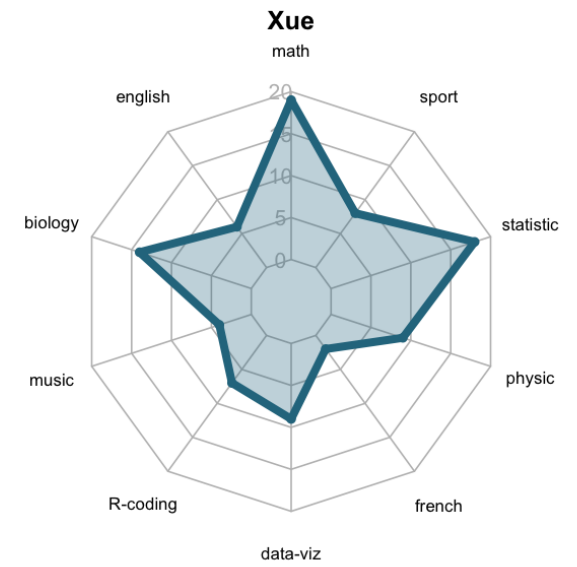
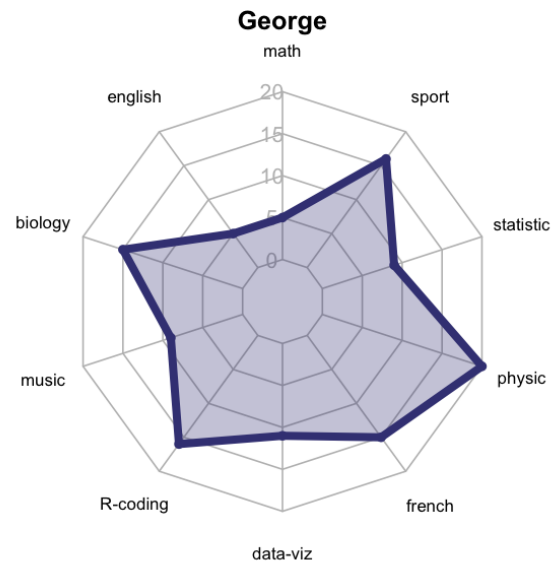
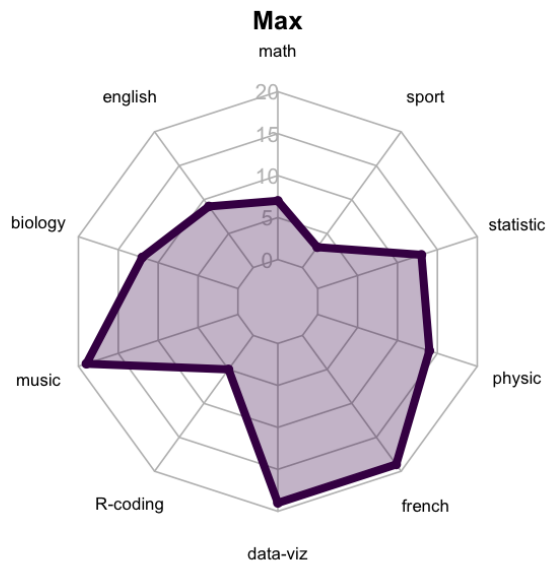
Multivariate: Star/Radar plots



n=21



n=7



- Univariate:
 - bar charts
 - histogram
 - box plot
 - ...
- Bivariate
 - clustered bar chart
 - stacked bar chart
 - 100% stacked bar chart
 - scatter plot
 - ...
- Tri-variate
 - scatter plot matrix
 - heat map
 - mosaic plot
 - ...
- Multivariate
 - parallel co-ordinates
 - SPLOM
 - ... and other techniques from a later lecture

Depicting Quantitative Data