

Data visualisation in R

Why is data visualisation important?

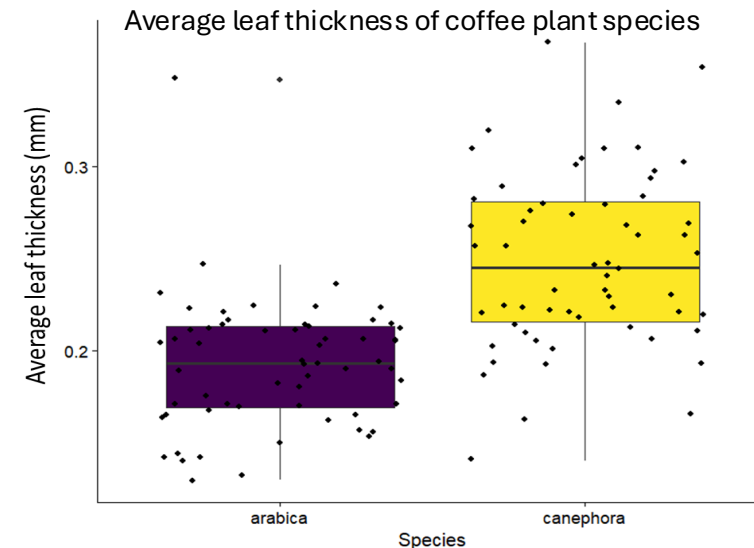
What is a good plot?

Which tools can I use?

Why is data visualisation important?

- Goal of an experiment is to **test one or several predictions** → with plots we can answer the initial hypotheses
- Transform measurements from generic numbers to shapes
- Plots enable us to visualise **patterns** and **trends** in our data
- Plots help us **communicate** scientific findings more effectively

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Image	Median_Red	Median_Green	Median_Blue	Median_Red	Median_Green	Median_Blue	SD_Red	SD_Green	SD_Blue	H_mean	S_mean	V_mean	H_mean	S_mean	V_mean	Time
2	20211127	43	86	9	49,26945	93,0762	28,23936	21,93391	25,99584	39,76566	93,50049	0,895349	0,337255	100,5388	0,6966	0,365005	1
3	20240730	131	175	2	143,6746	181,2805	39,11947	43,51777	31,53851	69,71103	75,26012	0,988571	0,686275	75,9056	0,784324	0,713331	2
4	20240730	132	176	2	144,46	181,8154	39,15565	43,41407	31,37476	69,74956	75,17241	0,988636	0,690196	75,71096	0,784641	0,713001	3
5	20240730	132	176	2	144,1477	182,122	39,2069	43,41938	31,25965	69,74268	75,17241	0,988636	0,690196	75,94275	0,784722	0,714204	4
6	20240730	133	176	2	145,1829	182,178	39,56299	43,29818	31,19448	70,0839	74,82759	0,988636	0,690196	75,56433	0,782833	0,714423	5
7	20240730	133	176	2	145,2215	182,2787	39,59996	43,24818	31,17419	69,92918	74,82759	0,988636	0,690196	75,51531	0,784067	0,714818	6
8	20240730	132	177	2	144,8316	182,5057	39,23964	43,25893	31,13209	69,83741	75,42857	0,988701	0,694118	75,77793	0,784995	0,715708	7
9	20240730	133	177	2	145,1989	182,5043	39,59373	43,30094	31,10708	70,17867	75,08571	0,988701	0,694118	75,66238	0,783053	0,715703	8
10	20240730	133	177	2	144,8439	182,6687	39,57343	43,30812	31,07079	70,18511	75,08571	0,988701	0,694118	75,85997	0,78336	0,716348	9
11	20240730	132	177	2	144,525	182,9137	39,42302	43,3416	31,03018	70,07063	75,42857	0,988701	0,694118	75,05206	0,784472	0,717308	10
12	20240730	134	177	2	145,6289	182,8899	39,43154	43,26797	31,03465	70,08376	74,74286	0,988701	0,694118	75,58405	0,784397	0,717215	11
13	20240730	133	177	2	145,5356	182,9826	39,28973	43,3397	31,06151	70,10831	75,08571	0,988701	0,694118	75,63626	0,785282	0,717579	12
14	20240730	132	177	2	144,5272	183,0926	39,03402	43,48571	31,08891	69,9765	75,42857	0,988701	0,694118	76,06239	0,786807	0,71801	13
15	20240730	133	177	2	145,0598	182,8386	39,11524	43,51678	31,17499	70,16338	75,08571	0,988701	0,694118	75,77397	0,786067	0,717014	14
16	20240730	132	177	2	144,3164	182,7637	38,85655	43,61792	31,23992	70,02438	75,42857	0,988701	0,694118	76,03006	0,787395	0,716721	15
17	20240730	132	177	2	144,2841	182,7293	38,73513	43,67653	31,25992	69,96238	75,42857	0,988701	0,694118	76,01946	0,788019	0,716585	16
18	20240730	133	177	2	144,9314	182,6233	38,83938	43,65261	31,26642	70,05304	75,08571	0,988701	0,694118	75,72855	0,787325	0,71617	17
19	20240730	133	177	2	144,444	182,7743	38,71276	43,68988	31,23783	69,92883	75,42857	0,988701	0,694118	75,96416	0,788184	0,716762	18
20	20240730	133	177	2	144,9994	182,7966	38,85198	43,61844	31,19862	70,0231	75,08571	0,988701	0,694118	75,7549	0,787458	0,716849	19
21	20240730	133	177	2	145,1021	182,87	38,83585	43,55968	31,16759	69,98492	75,08571	0,988701	0,694118	75,7329	0,787631	0,717137	20
22	20240730	133	177	2	144,9192	183,0421	38,80454	43,60816	31,12222	69,9373	75,08571	0,988701	0,694118	75,85837	0,788002	0,717812	21
23	20240730	133	177	2	145,0022	183,094	38,83312	43,52525	31,11943	69,95921	75,08571	0,988701	0,694118	75,84288	0,787906	0,718016	22
24	20240730	133	177	2	145,2235	183,0937	38,97364	43,54601	31,09591	70,09211	75,08571	0,988701	0,694118	75,76607	0,787138	0,718014	23
25	20240730	133	177	2	145,1151	183,1719	38,95108	43,56398	31,09824	70,06402	75,08571	0,988701	0,694118	75,83275	0,787352	0,718321	24
26	20240730	133	177	2	145,2098	183,1726	39,06955	43,5317	31,09231	70,18366	75,08571	0,988701	0,694118	75,80652	0,786706	0,718324	25
27	20240730	133	177	2	145,0844	183,207	39,09465	43,54586	31,0914	70,21795	75,08571	0,988701	0,694118	75,87201	0,786609	0,718459	26
28	20240730	133	177	2	145,3799	183,211	39,13809	43,53691	31,08591	70,26091	75,08571	0,988701	0,694118	75,75499	0,786377	0,718475	27
29	20240730	132	177	2	144,8872	183,3217	39,04719	43,58075	31,0839	70,20418	75,42857	0,988701	0,694118	75,98391	0,787002	0,718909	28
30	20240730	132	177	2	144,977	183,3542	38,91243	43,57366	31,08278	70,09445	75,42857	0,988701	0,694118	75,9416	0,787775	0,719036	29
31	20240730	134	177	2	145,0774	183,2439	39,09047	43,50585	31,0949	70,18017	75,44007	0,988701	0,694118	75,54407	0,787117	0,718604	30
32	20240730	134	177	2	146,166	183,2285	39,09522	43,50339	31,08136	70,25745	75,42857	0,988701	0,694118	75,42845	0,786631	0,718543	31
33	20240730	133	177	2	145,4949	183,3734	38,95448	43,54549	31,07331	70,1799	75,08571	0,988701	0,694118	75,73692	0,787567	0,719111	32
34	20240730	133	177	2	145,5726	183,357	38,98257	43,55222	31,06697	70,20758	75,08571	0,988701	0,694118	75,70266	0,787395	0,719047	33
35	20240730	133	178	2	145,7626	183,4186	38,82303	43,52424	31,06489	70,13341	75,34091	0,988764	0,698039	75,63619	0,787791	0,719288	34
36	20240730	132	178	2	144,5731	183,5113	39,36338	43,59356	31,04617	70,52261	75,68182	0,988764	0,698039	76,20761	0,785499	0,719652	35
37	20240730	133	178	2	145,1505	183,4727	39,24451	43,5723	31,05067	70,42758	75,34091	0,988764	0,698039	75,94229	0,786102	0,719501	36
38	20240730	133	178	2	145,1001	183,5886	38,97465	43,54558	31,04354	70,20781	75,34091	0,988764	0,698039	75,96877	0,787707	0,719955	37
39	20240730	134	178	2	145,819	183,5366	39,15151	43,5131	31,02912	70,35957	75	0,988764	0,698039	75,65754	0,786603	0,719752	38
40	20240730	132	178	2	144,8968	183,6595	39,07877	43,55946	31,0247	70,30374	75,68182	0,988764	0,698039	76,08628	0,787222	0,720234	39
41	20240730	134	178	2	145,8523	183,5744	39,23078	43,50975	31,0307	70,4332	75	0,988764	0,698039	75,68012	0,786295	0,7199	40
42	20240730	134	178	2	146,144	183,5559	39,25107	43,49635	31,02843	70,45507	75	0,988764	0,698039	75,55336	0,786163	0,719827	41
43	20240730	134	178	2	145,8702	183,5682	39,44479	43,50964	31,02494	70,63854	75	0,988764	0,698039	75,69406	0,785122	0,719875	42
44	20240730	133	178	2	145,0444	183,7716	39,02859	43,56471	31,00815	70,30458	75,34091	0,988764	0,698039	76,05352	0,787625	0,720073	43

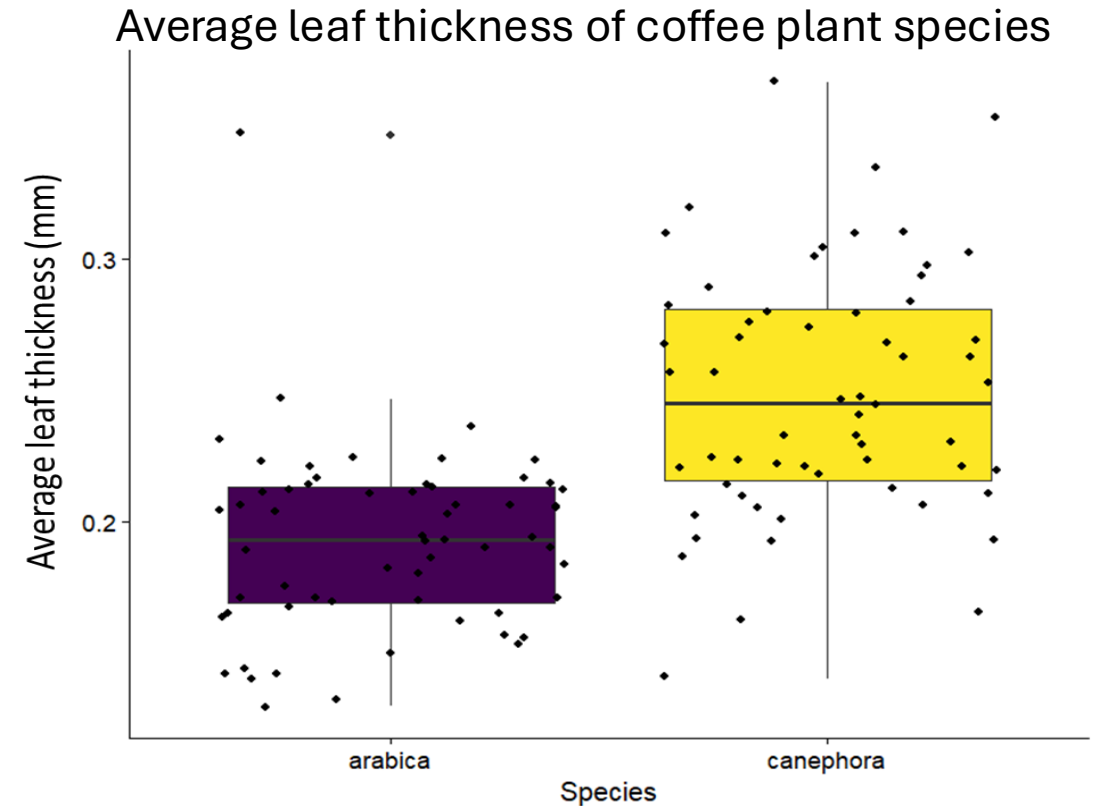


What is a good plot?

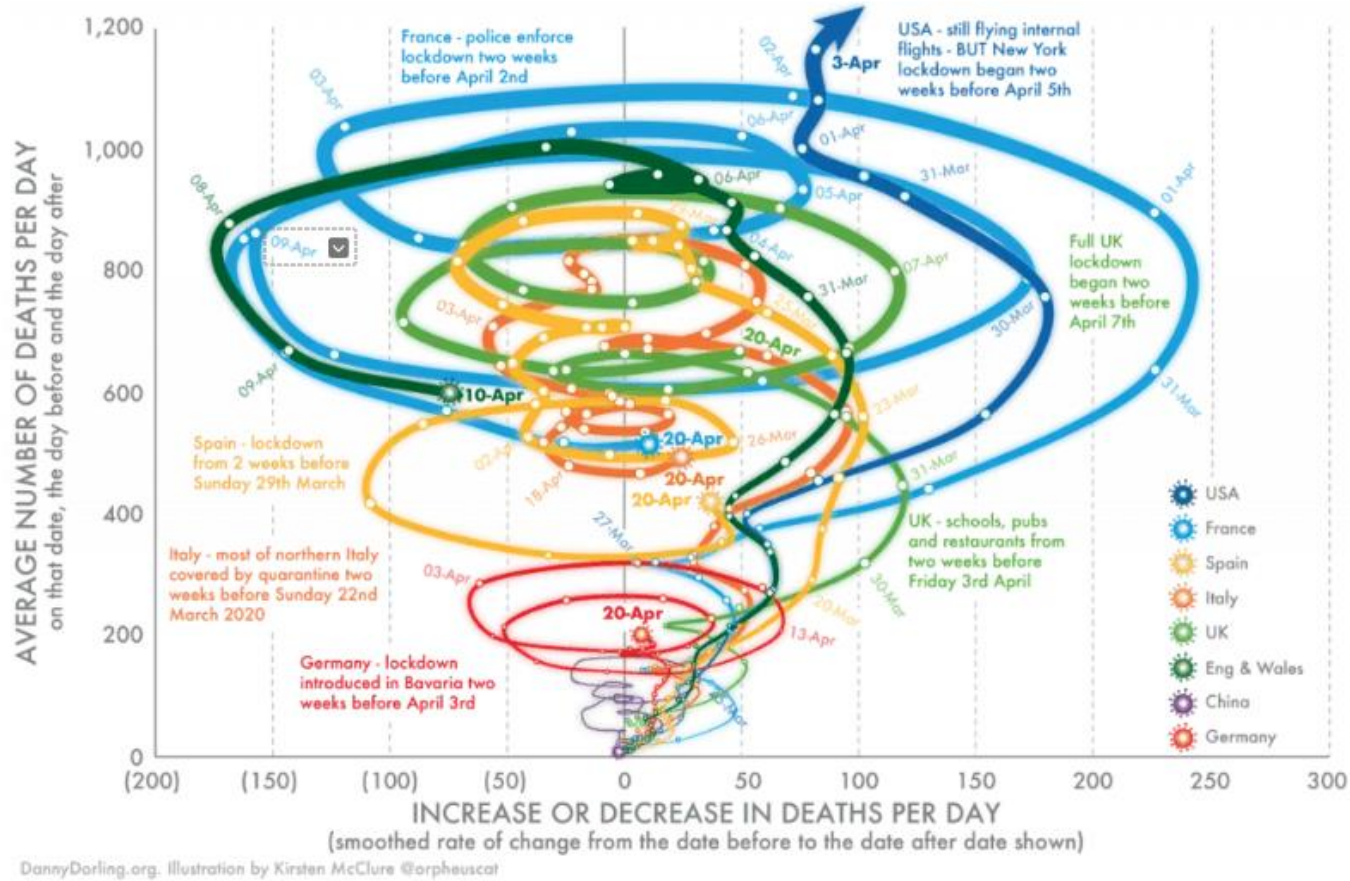
- Conveys relevant information
- Easy to understand
- Clear and visible
- Includes titles, legends, and units
- Use contrasting colours

Check these out

- <https://coolers.co/palettes/popular>
- <https://hclwizard.org>
- <https://www.datanovia.com/en/blog/the-a-z-of-r-colorbrewer-palette/>
- <https://bpb-us-e1.wpmucdn.com/sites.ucsc.edu/dist/d/276/files/2015/10/colorbynames.png?bid=276>

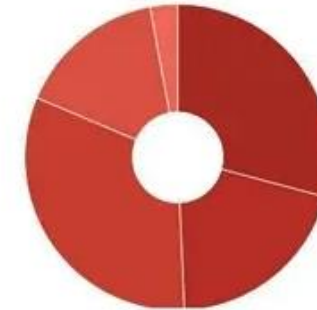


Examples of bad plots

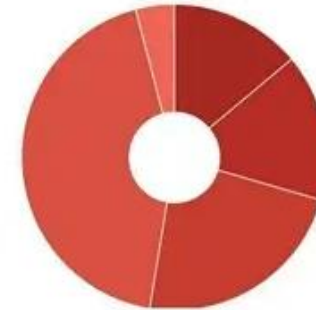


0-19 20-39 40-59 60-79 80+

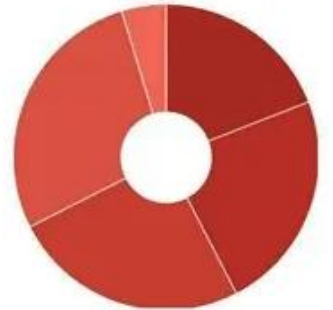
The Woodlands



Shenandoah



Oak Ridge North



Source: U.S. Census Bureau

What is a good plot?

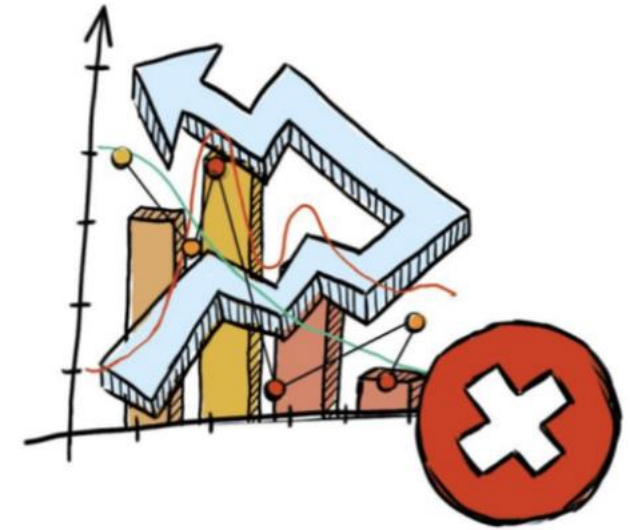
To create a good plot:

- Think about your hypotheses: what do you want to show?

→ **Relevance** is important!

- Think about your experimental design: what type of variables do you have?

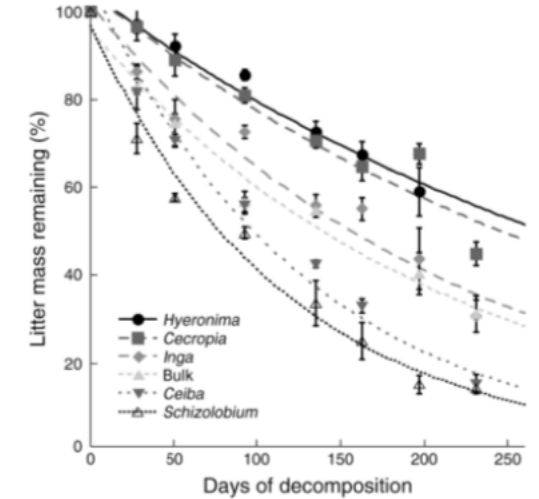
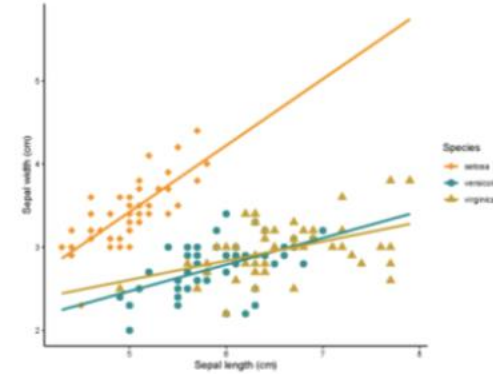
→ Choose the right type of plot



How to choose the appropriate plot?

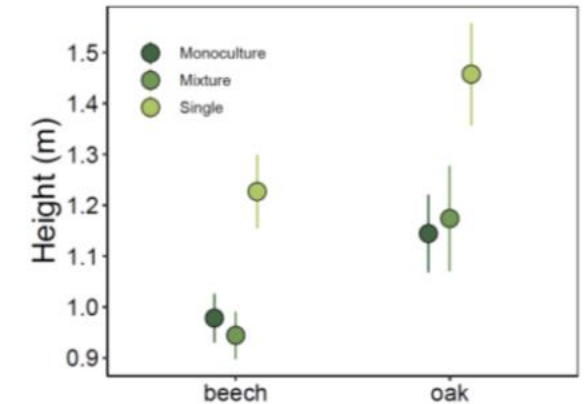
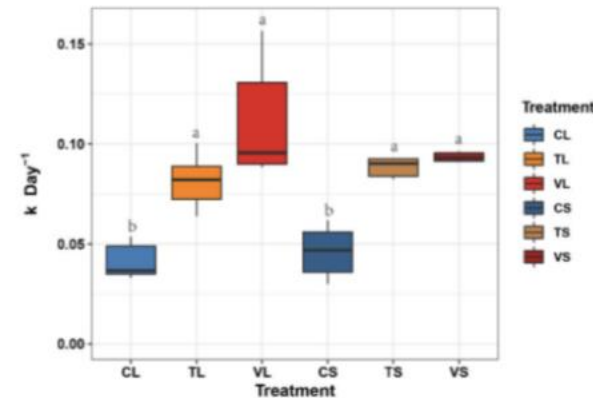
- Relationships: scatter plots, line plots

- Continuous variables
- Evolution over time



- Comparisons: box plots, bar plots, point plots

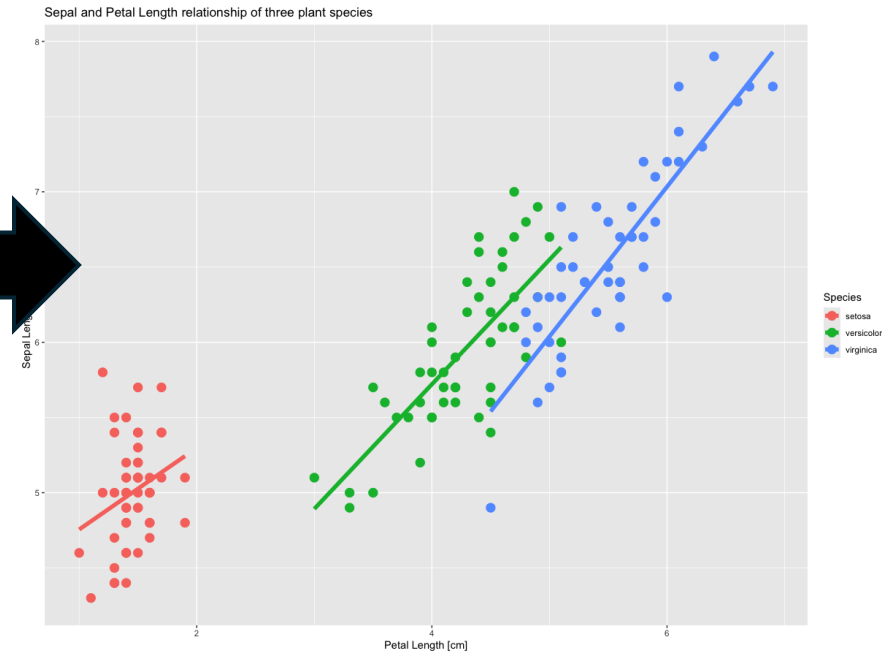
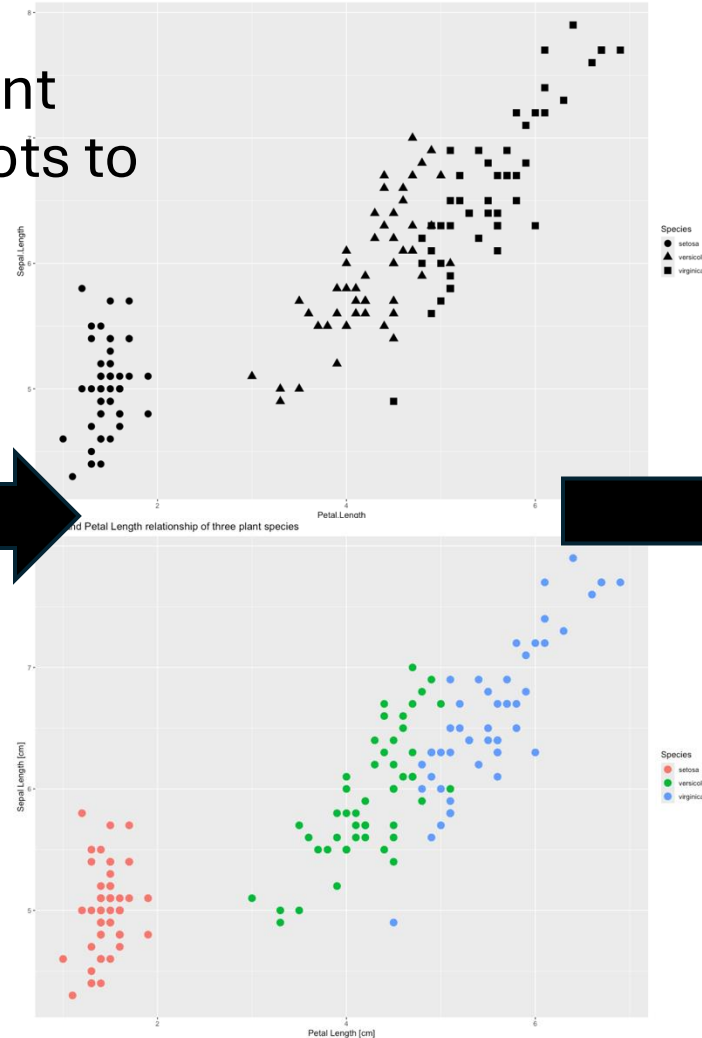
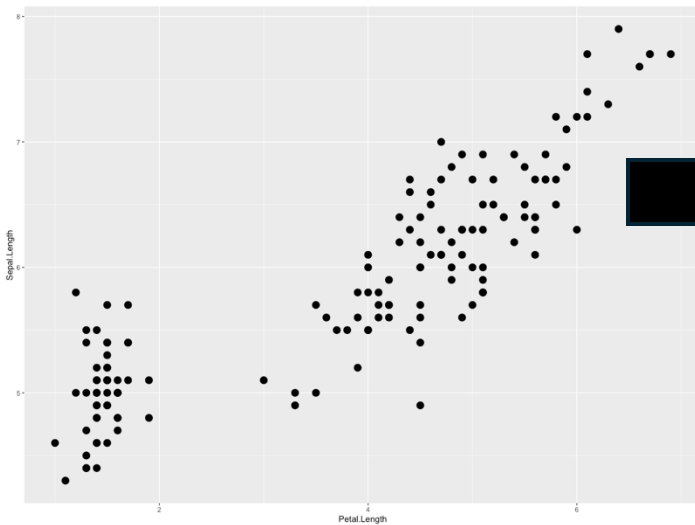
- Categorical variables
- Differences between groups



Choice of final plot

Explore your data, try different colours, shapes, types of plots to discover potential trends

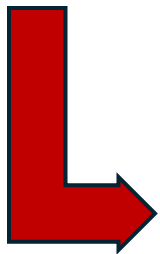
Final plot is obtained through trial and error!



Which tools can I use?



- In R, there are mainly two options to plot your data:
 - base R: built-in function that allow you to create different plots, easy to use for basic graphs
 - ggplot: R-package including a large variety of plots and esthetics (colours, shapes, facets,...)



Today we will mainly
be using ggplot!



Which tools can I use?

- Check the material on moodle:
 - Visualisation in R: examples of codes to create plots
 - Cheat Sheets: summaries of command for base R and ggplot
 - Visualisation tasks: small tasks for training
 - Your own data: you can already importing it on r and start exploring