

# DA-STATS

## Topic 03: Data Visualisation And Reporting Techniques

Shan E Ahmed Raza  
Department of Computer Science  
University of Warwick  
[shan.raza@warwick.ac.uk](mailto:shan.raza@warwick.ac.uk)



# Outline

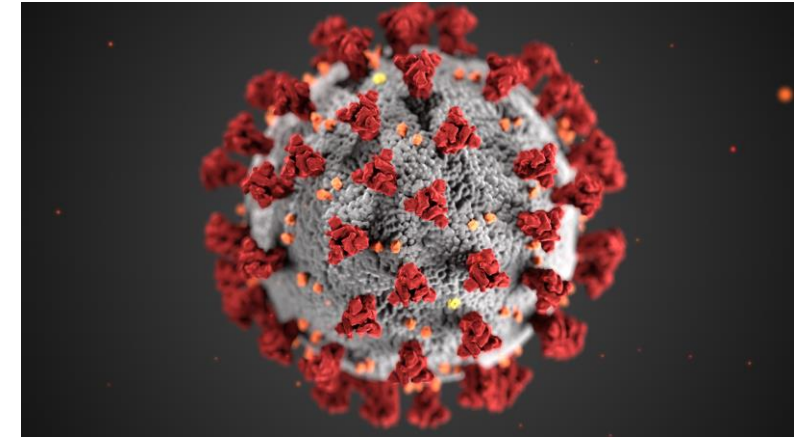


- The Purpose of Data Visualisation
- Appropriate Data Visualisation
- Choosing the Right Visualisation
- Various Visualisation Methods

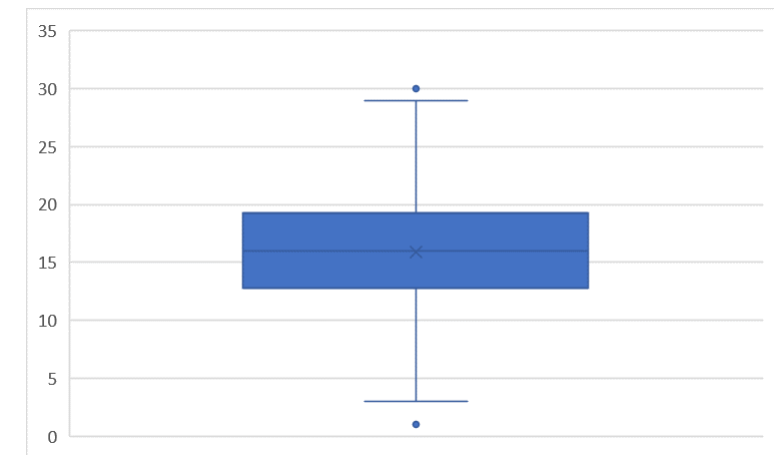
# The Purpose of Data Visualisation



- A picture is worth a thousand words
- Get insight
- Organisation

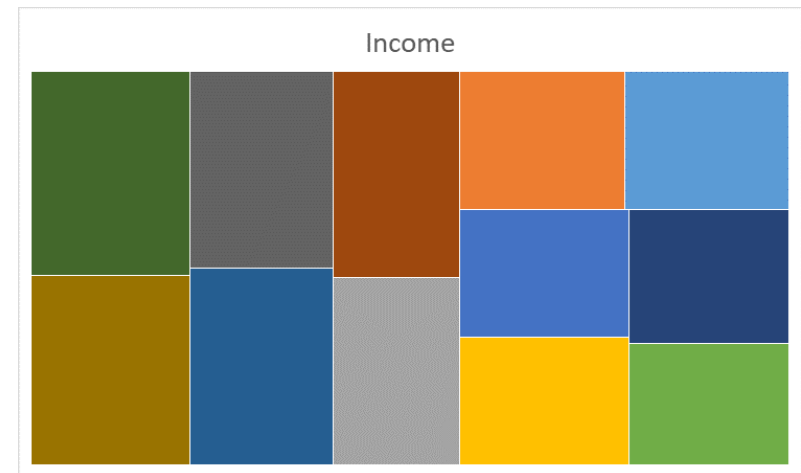
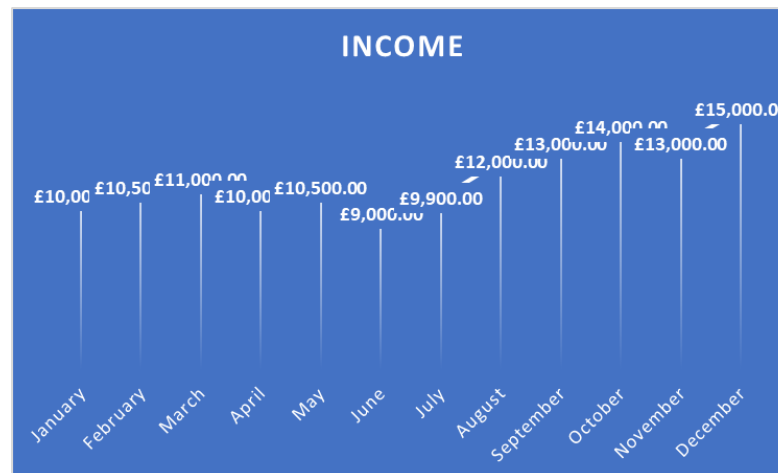
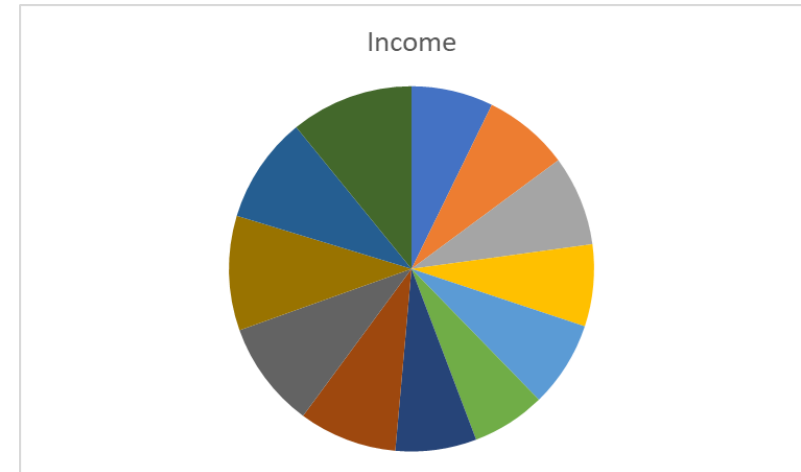
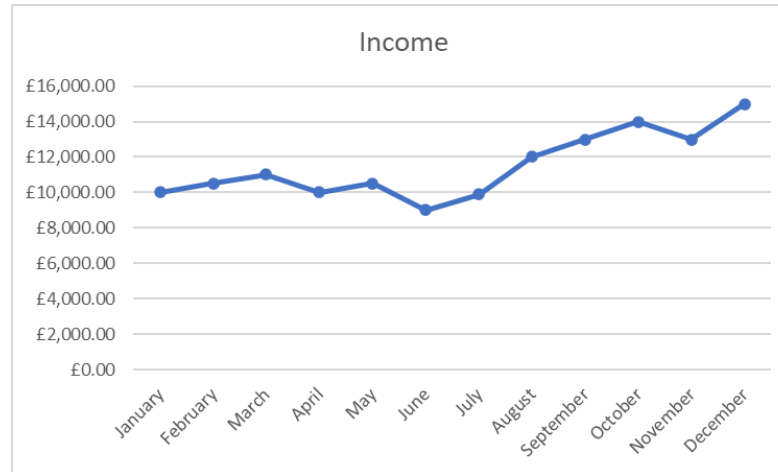


15	8	20	16	12	18	14	22	17	5
19	15	18	29	6	13	16	19	10	24
15	3	26	30	13	17	7	16	23	25
1	15	18	14	5	27	16	20	14	6
24	14	20	25	21	15	17	8	23	21
17	14	10	13	18	16	21	9	11	22
15	12	9	16	20	11	13	22	17	13
9	22	16	12	19	17	14	10	19	18
11	16	12	18	13	17	15	14	15	28



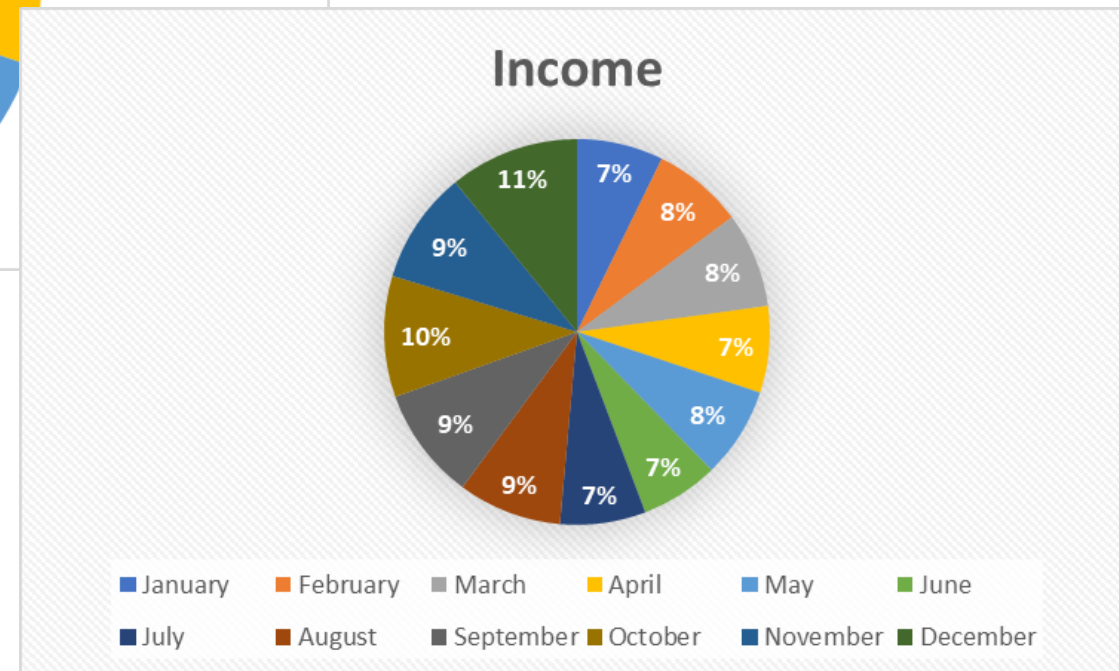
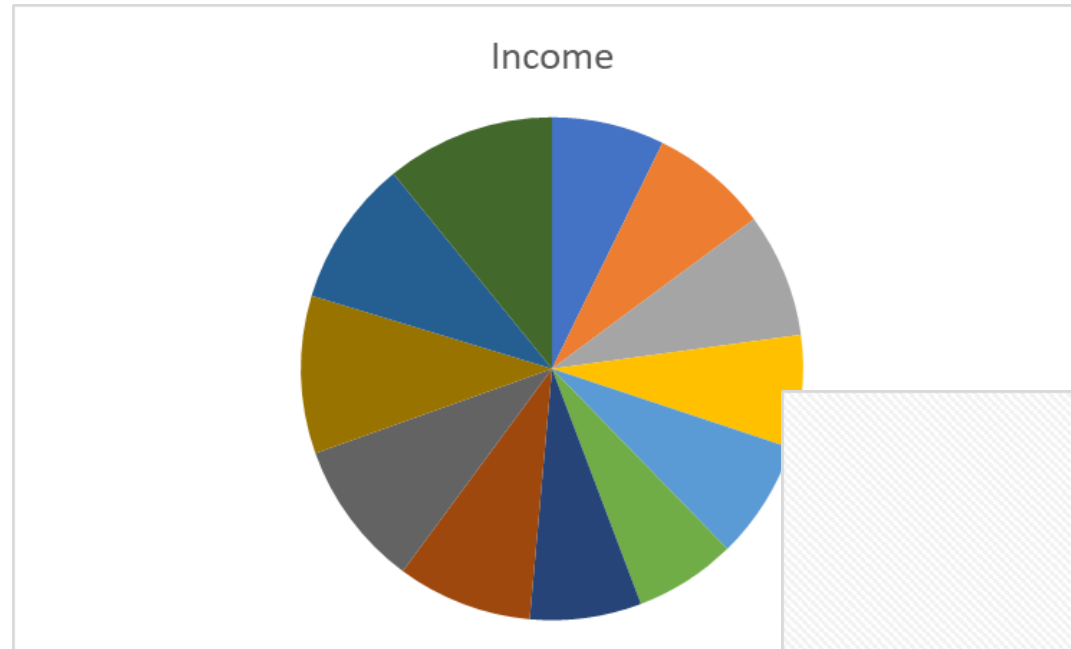
# Appropriate Data Visualisation

Month	Income
January	£10,000.00
February	£10,500.00
March	£11,000.00
April	£10,000.00
May	£10,500.00
June	£9,000.00
July	£9,900.00
August	£12,000.00
September	£13,000.00
October	£14,000.00
November	£13,000.00
December	£15,000.00



# Appropriate Data Visualisation

Month	Income
January	£10,000.00
February	£10,500.00
March	£11,000.00
April	£10,000.00
May	£10,500.00
June	£9,000.00
July	£9,900.00
August	£12,000.00
September	£13,000.00
October	£14,000.00
November	£13,000.00
December	£15,000.00



# Choosing the Right Data Visualisation



- What is the message?
- Who is the audience?
- What do you want to show?
  - Distribution
  - Comparison
  - Ranking
  - Relationship
  - Timeline/trends
  - Composition
- Can you present the data in different ways?

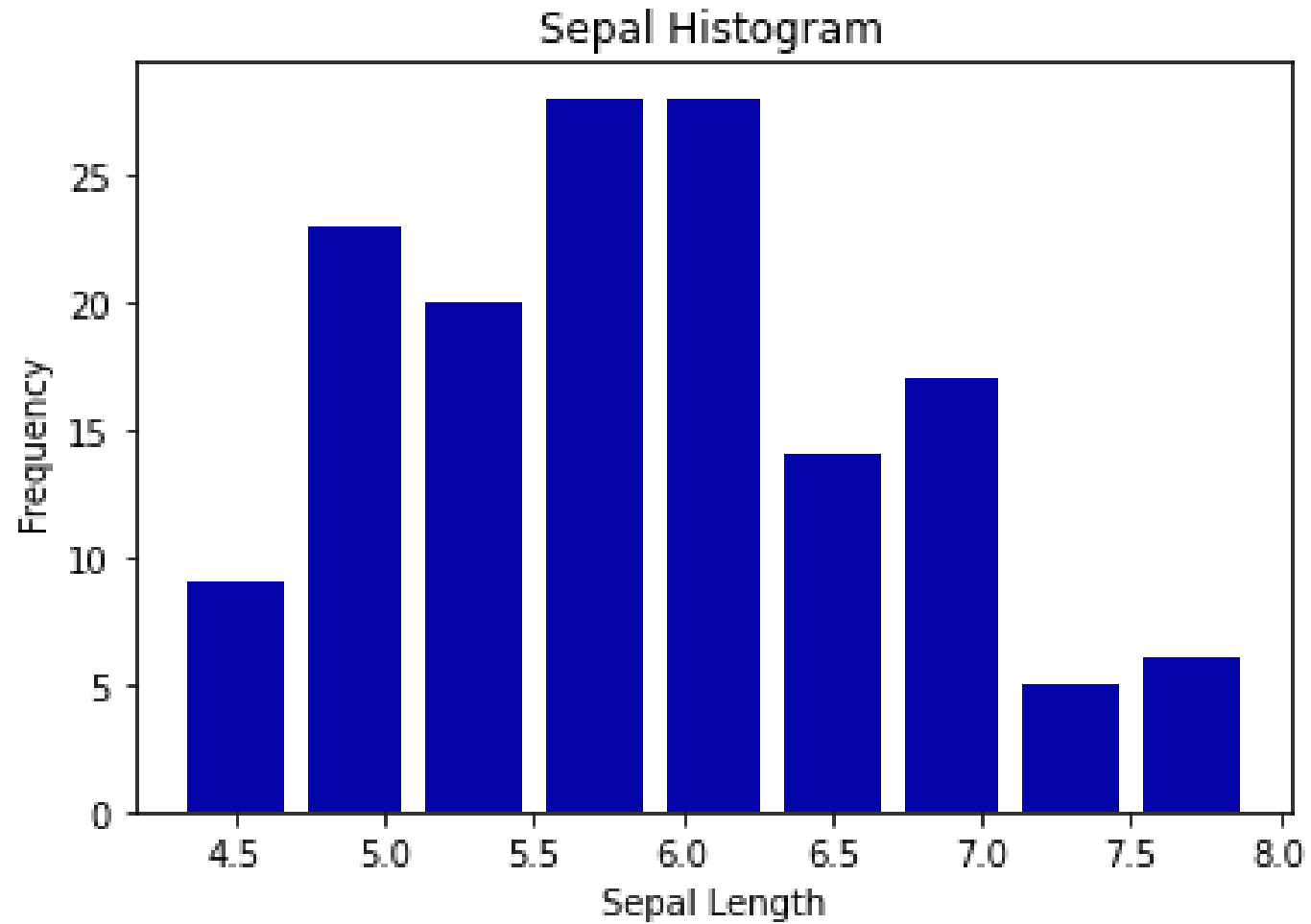
# Iris data set

- Three different species of iris
  - Iris-Setosa
  - Iris-Virginica
  - Iris-Versicolor



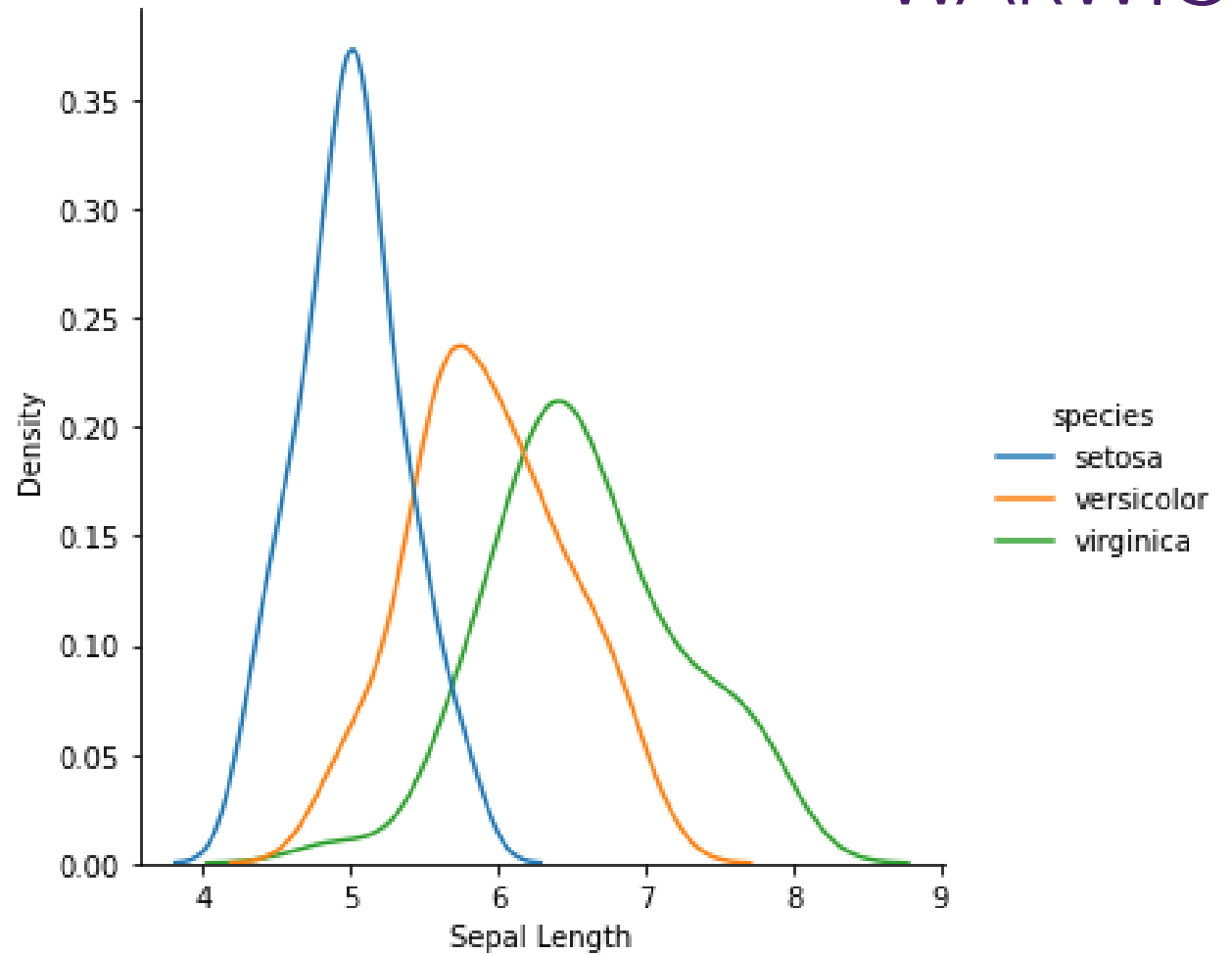
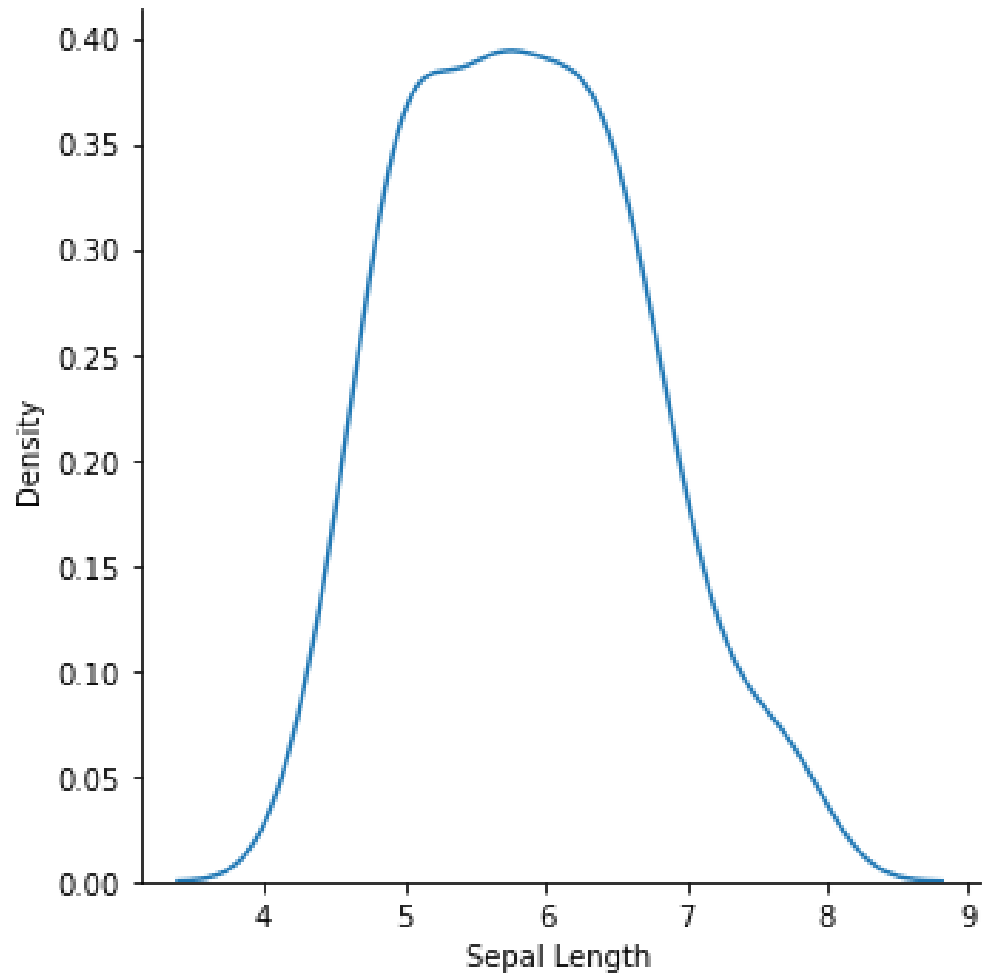
	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

# Histogram

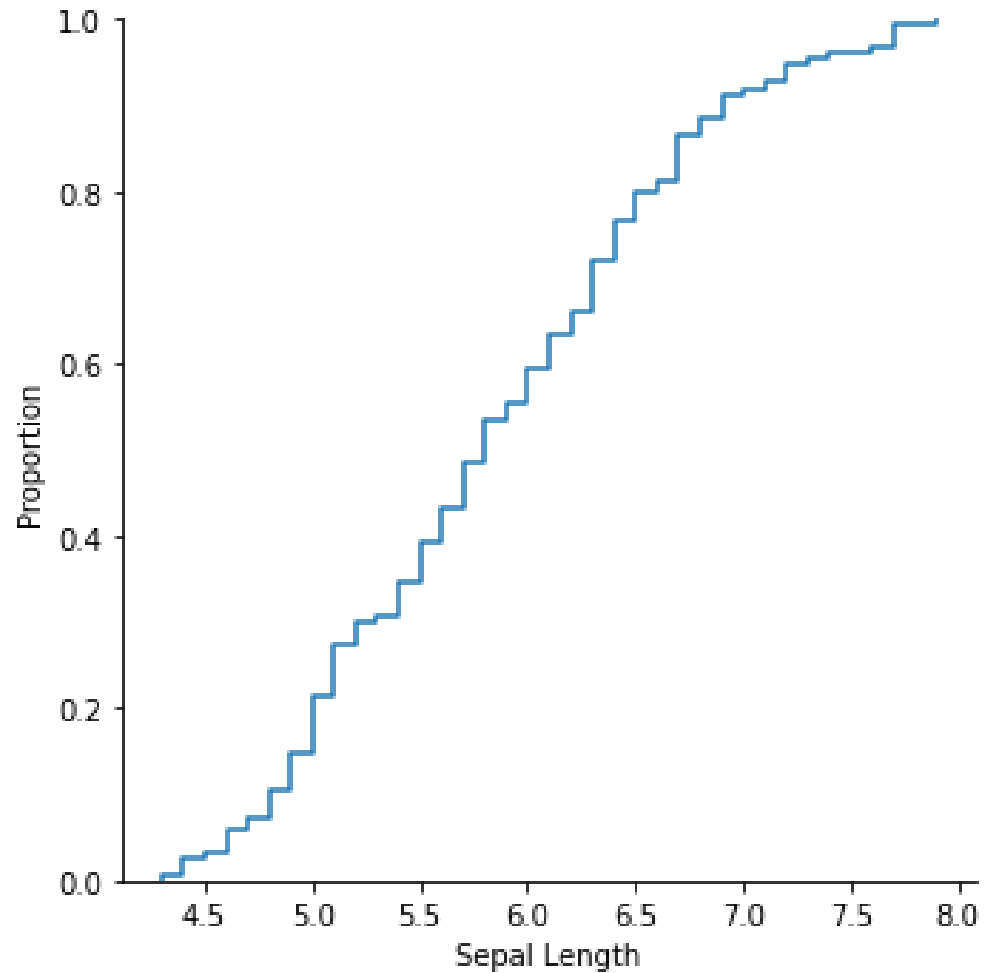




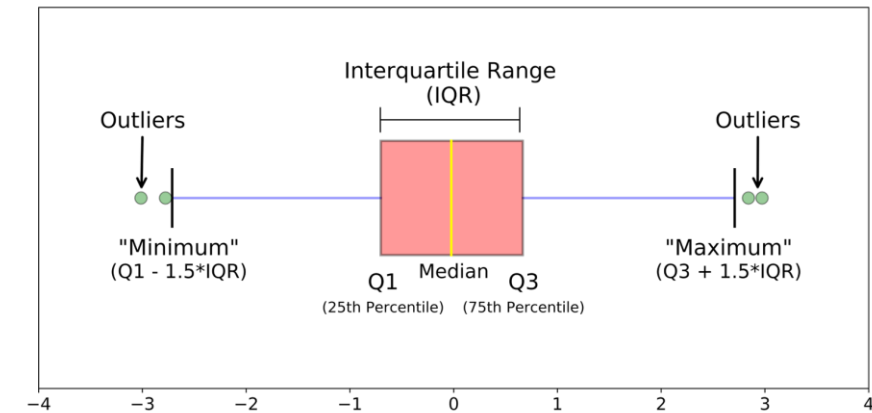
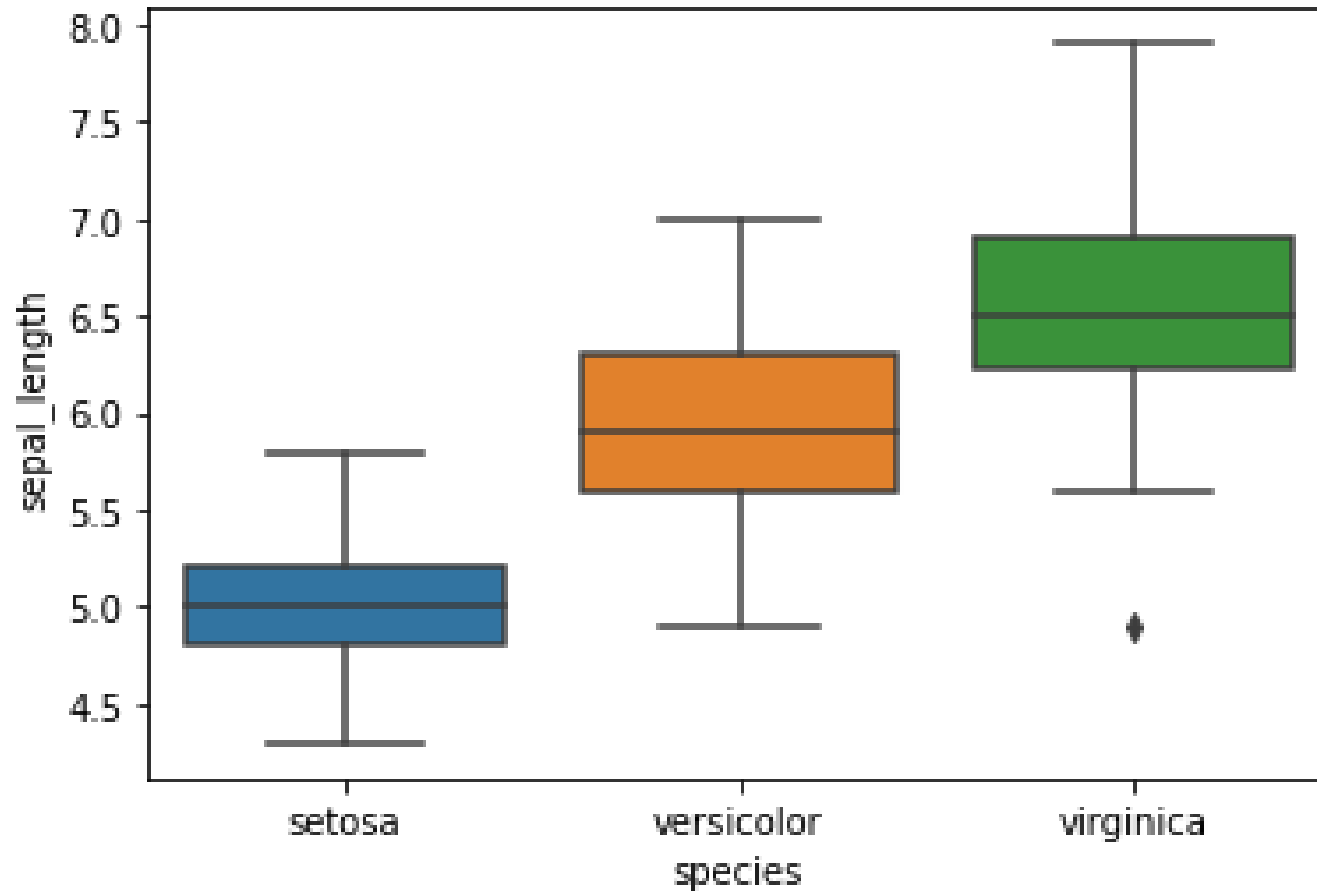
# Density Plot



# Cumulative Distribution

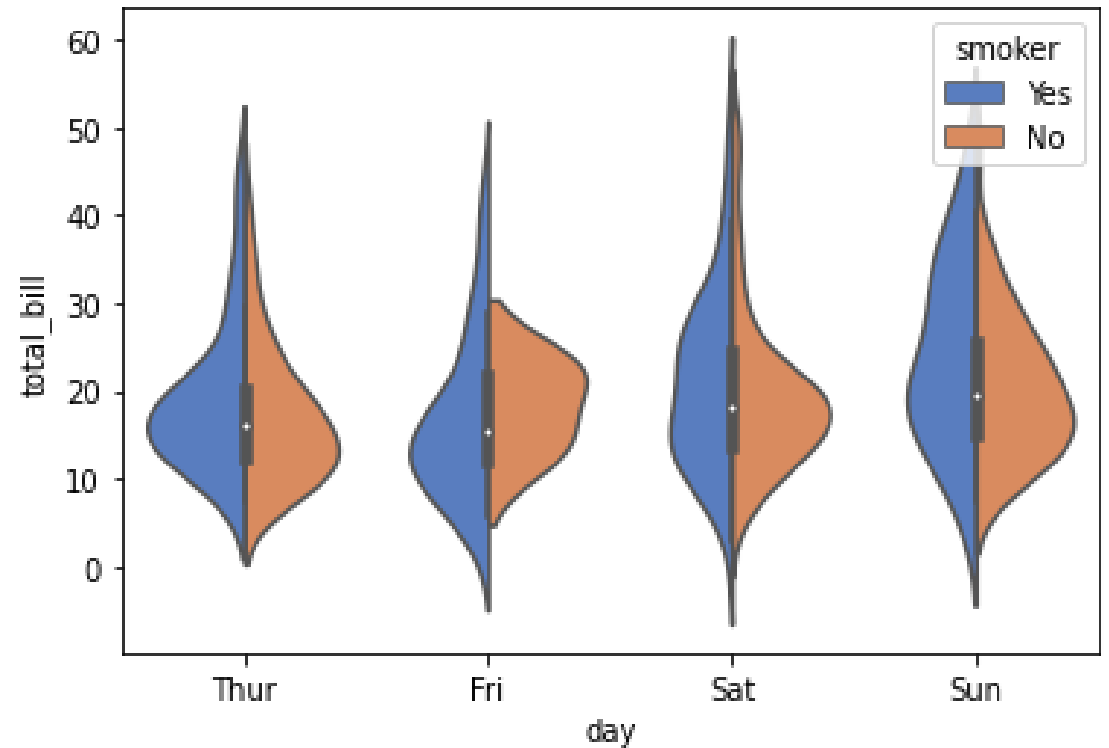
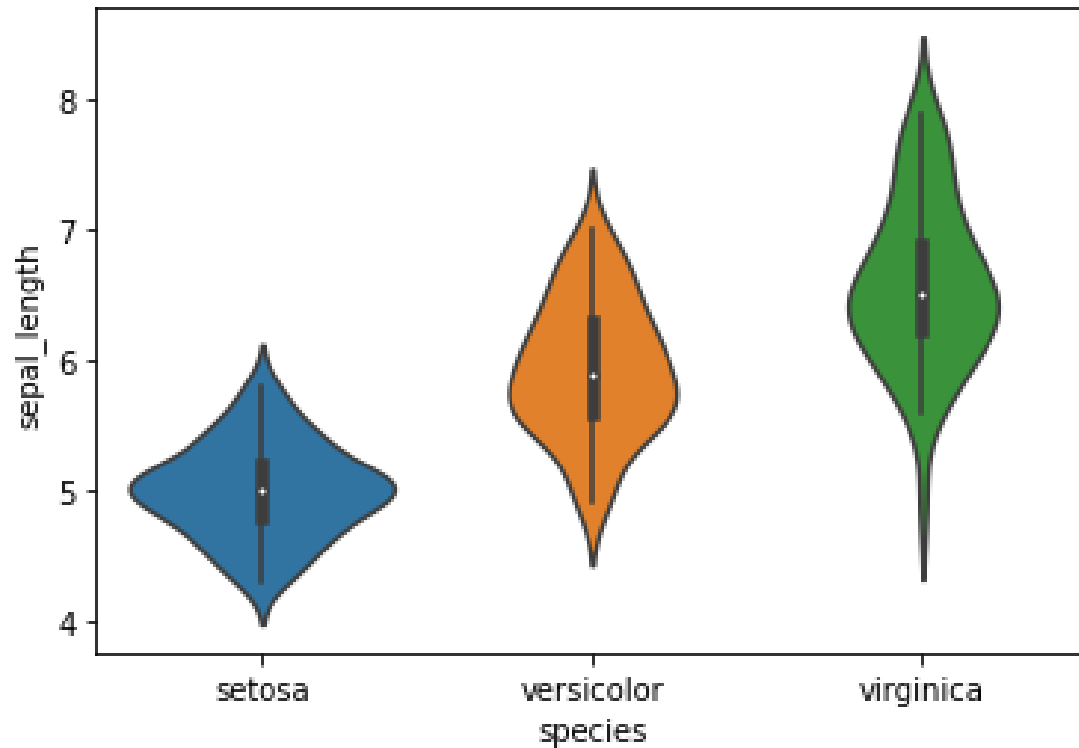
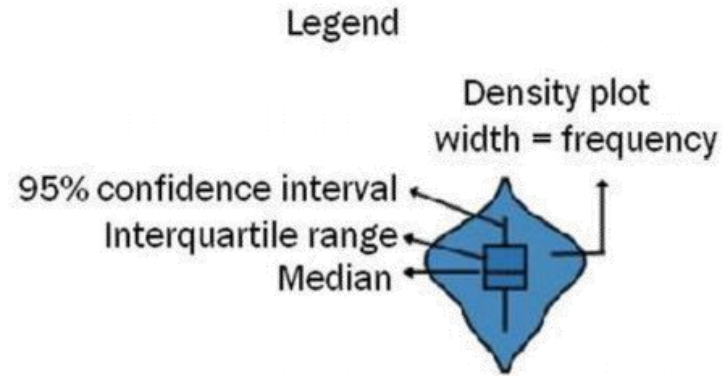


# Box (Box-Whisker) Plot

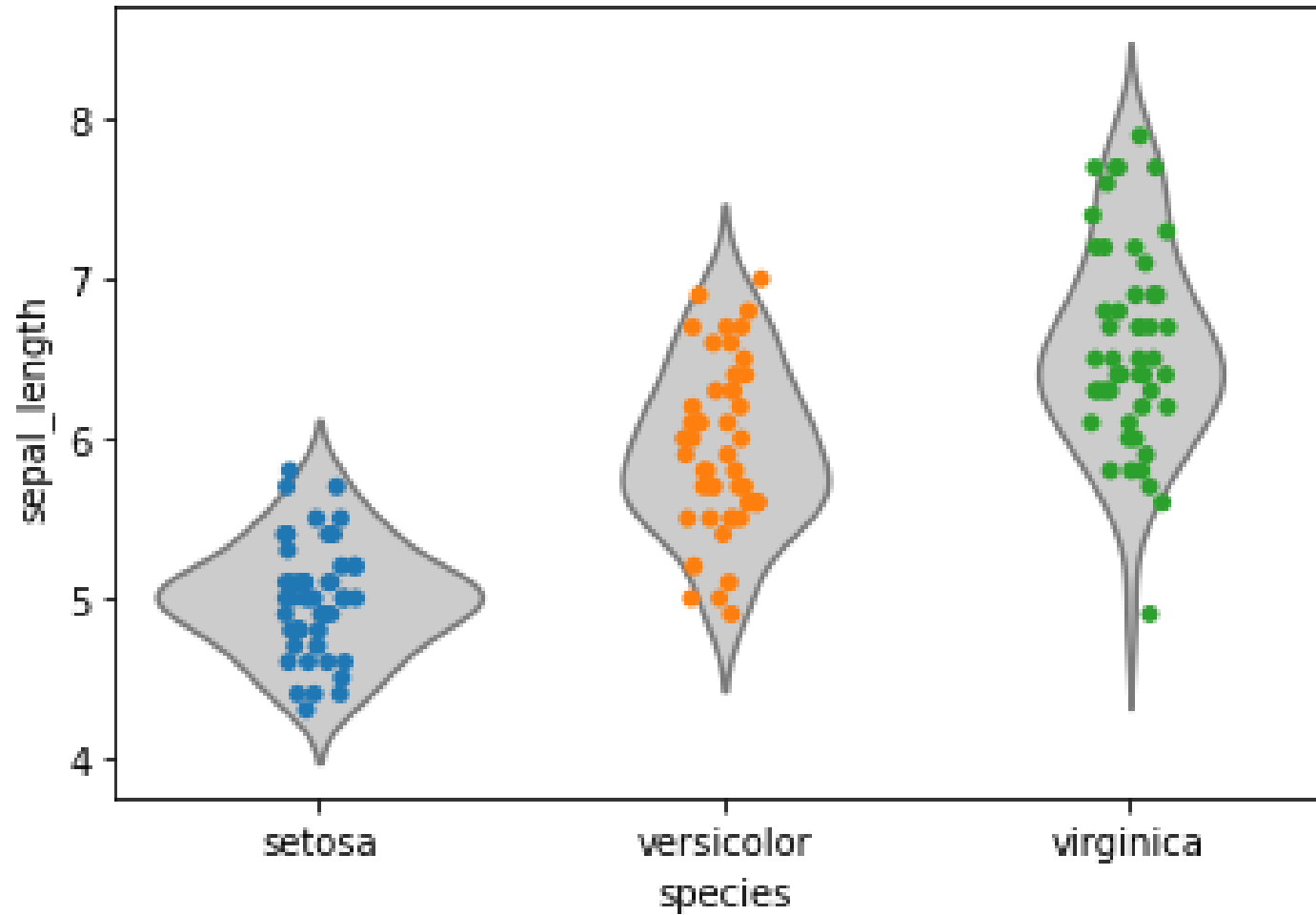


<https://towardsdatascience.com/understanding-boxplots-5e2df7bcdb51>

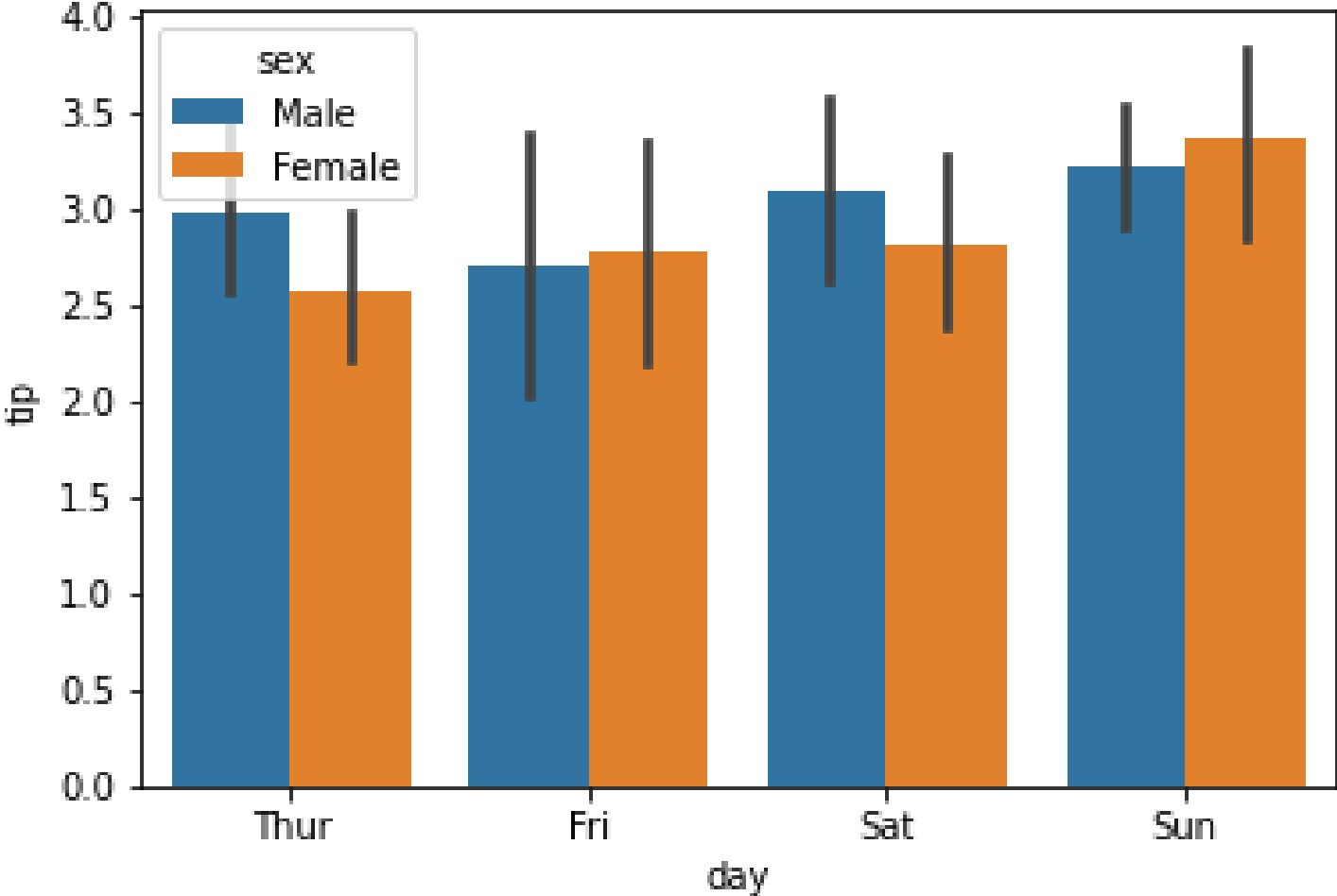
# Violin Plot



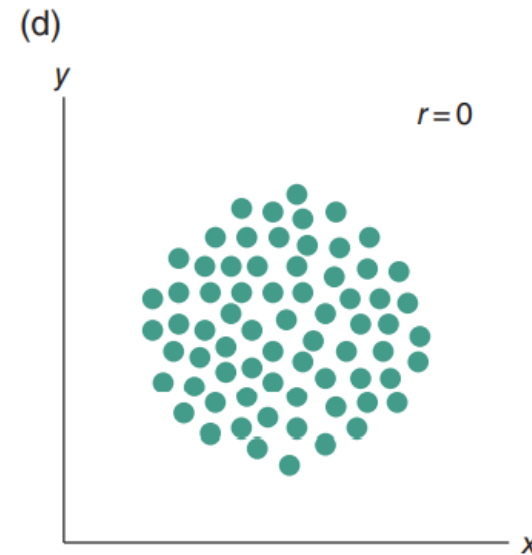
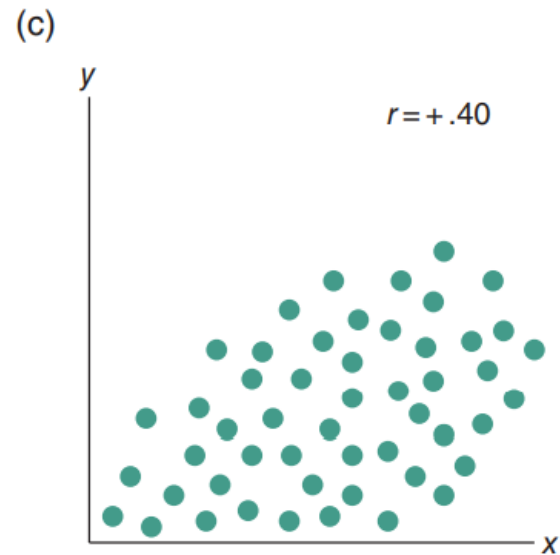
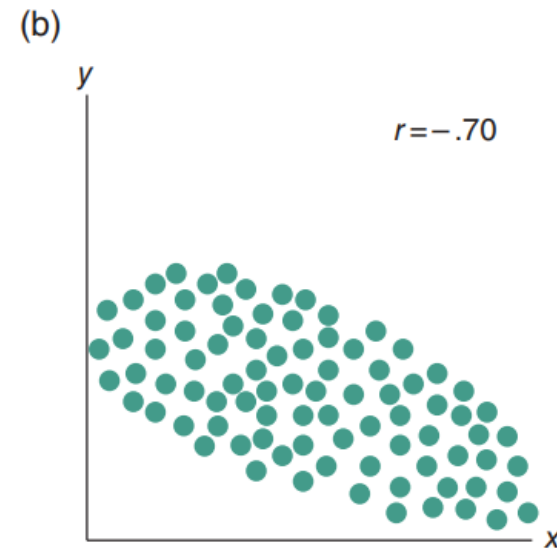
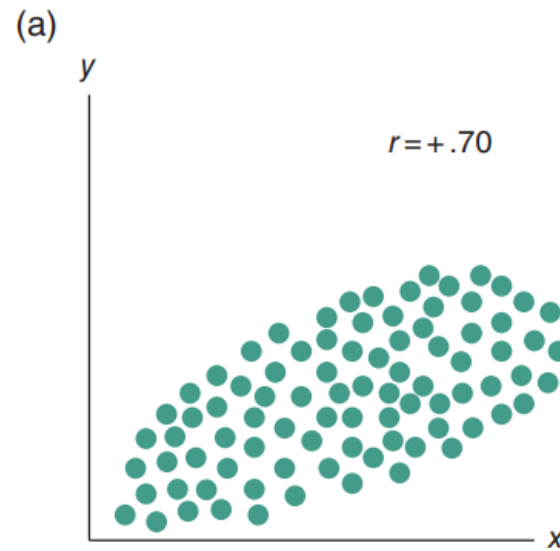
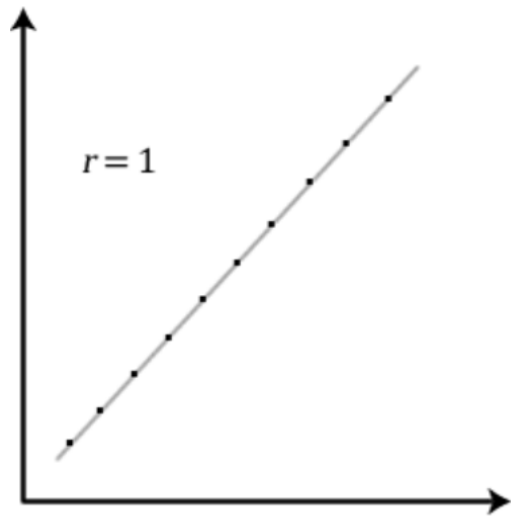
# Strip Plot



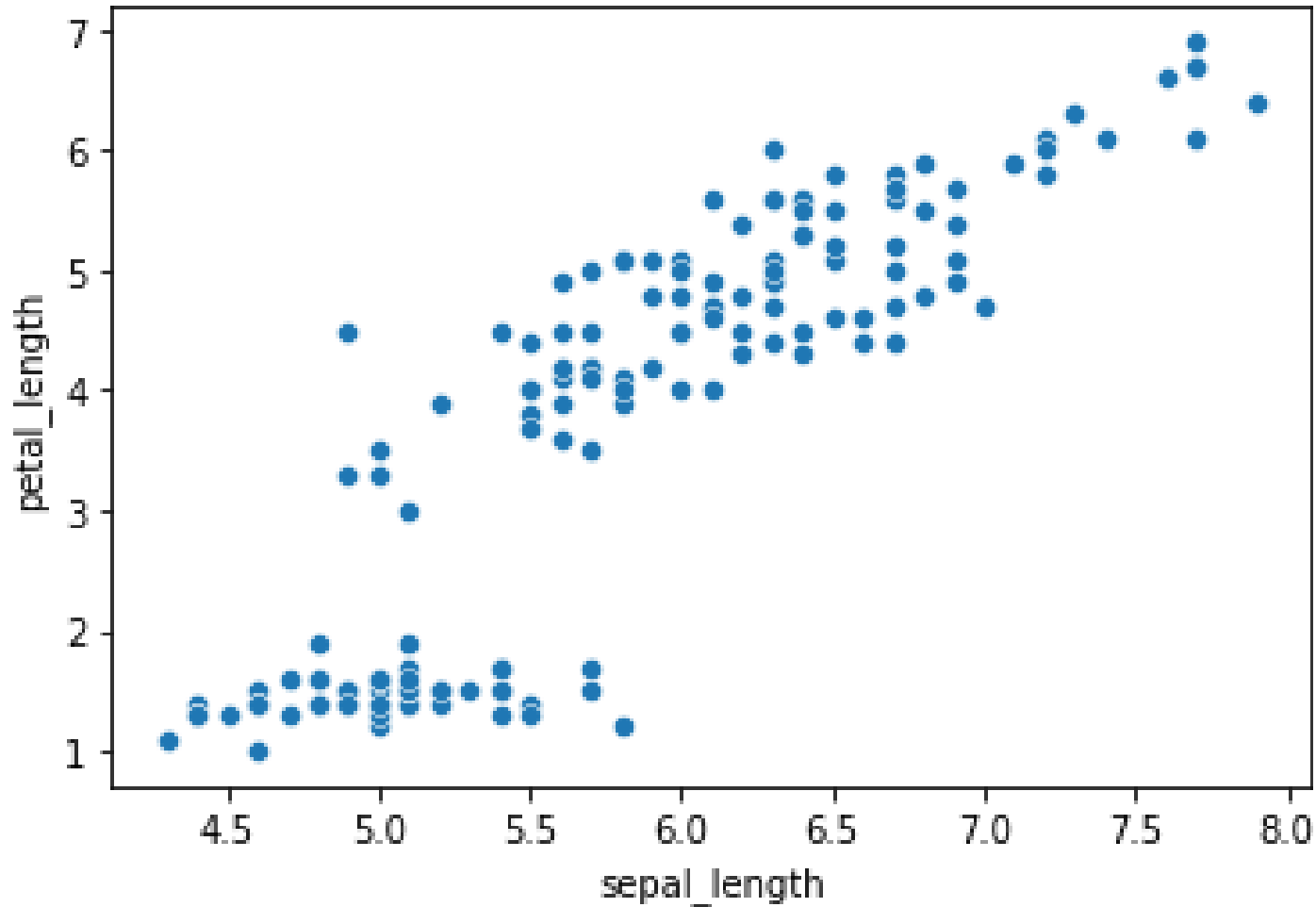
# Bar Chart



# Scatter Plot



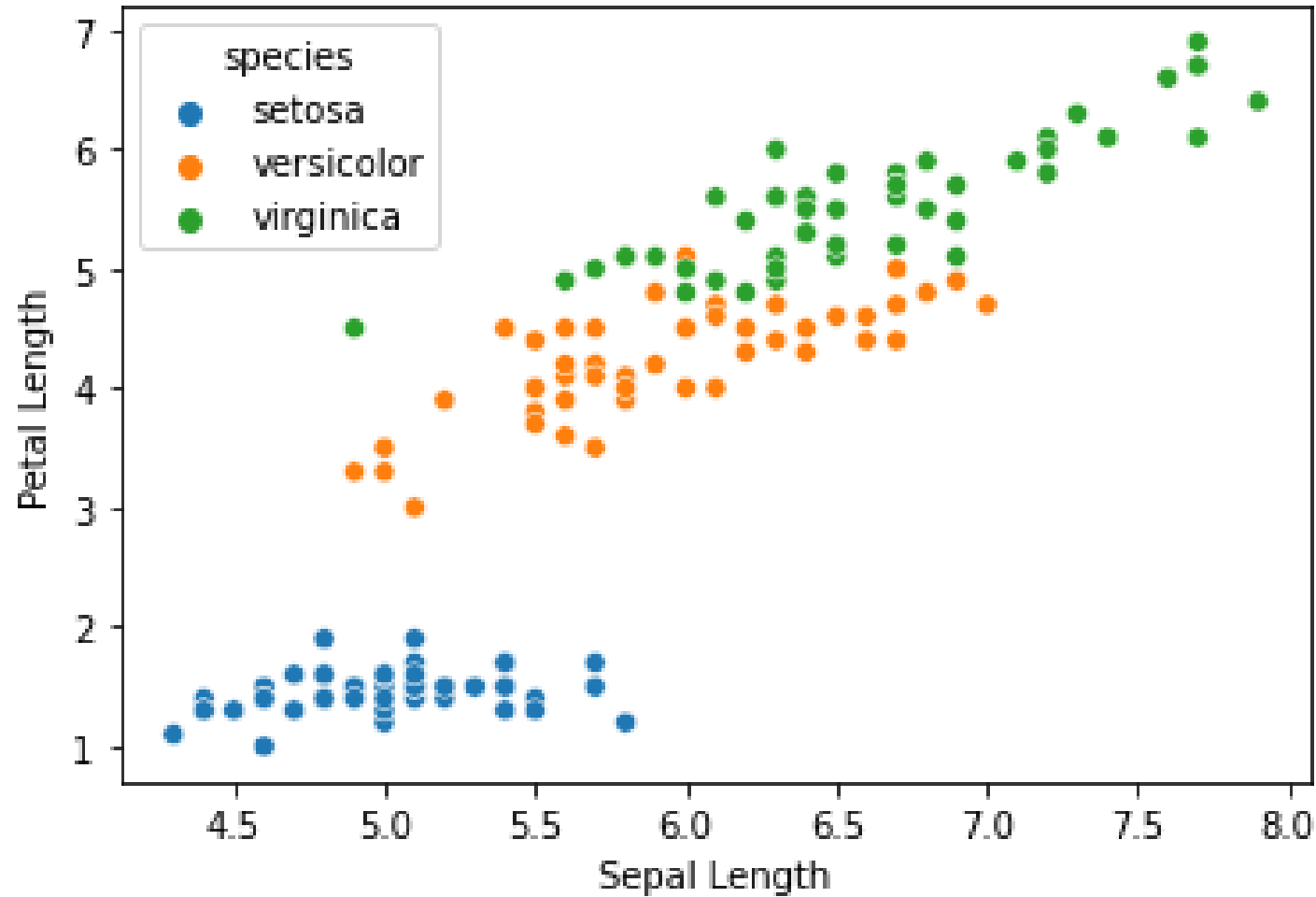
# Scatter Plot



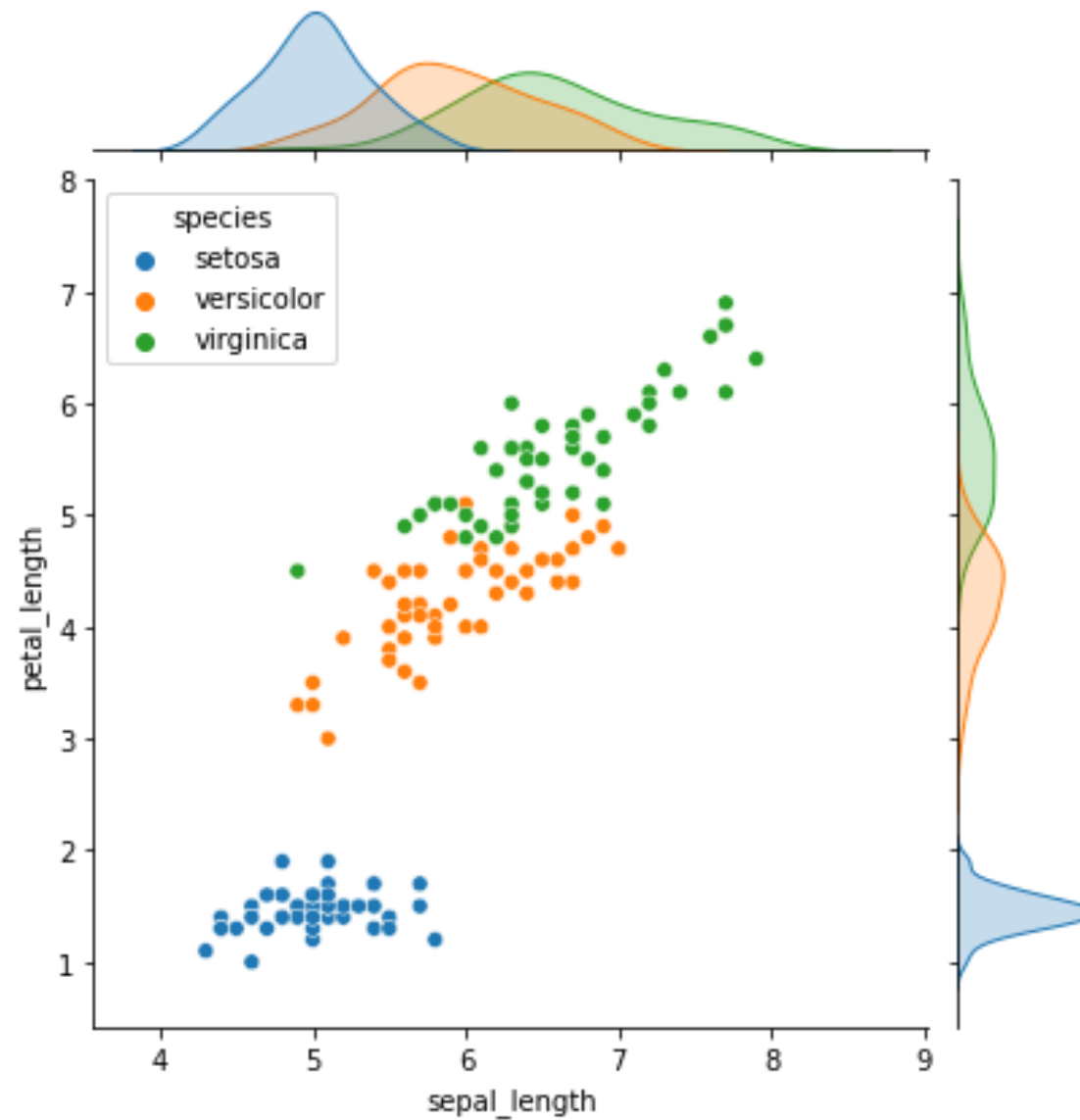
Pearson Correlation: 0.87  
Spearman Correlation: 0.88  
Kendall Tau: 0.71



# Scatter Plot

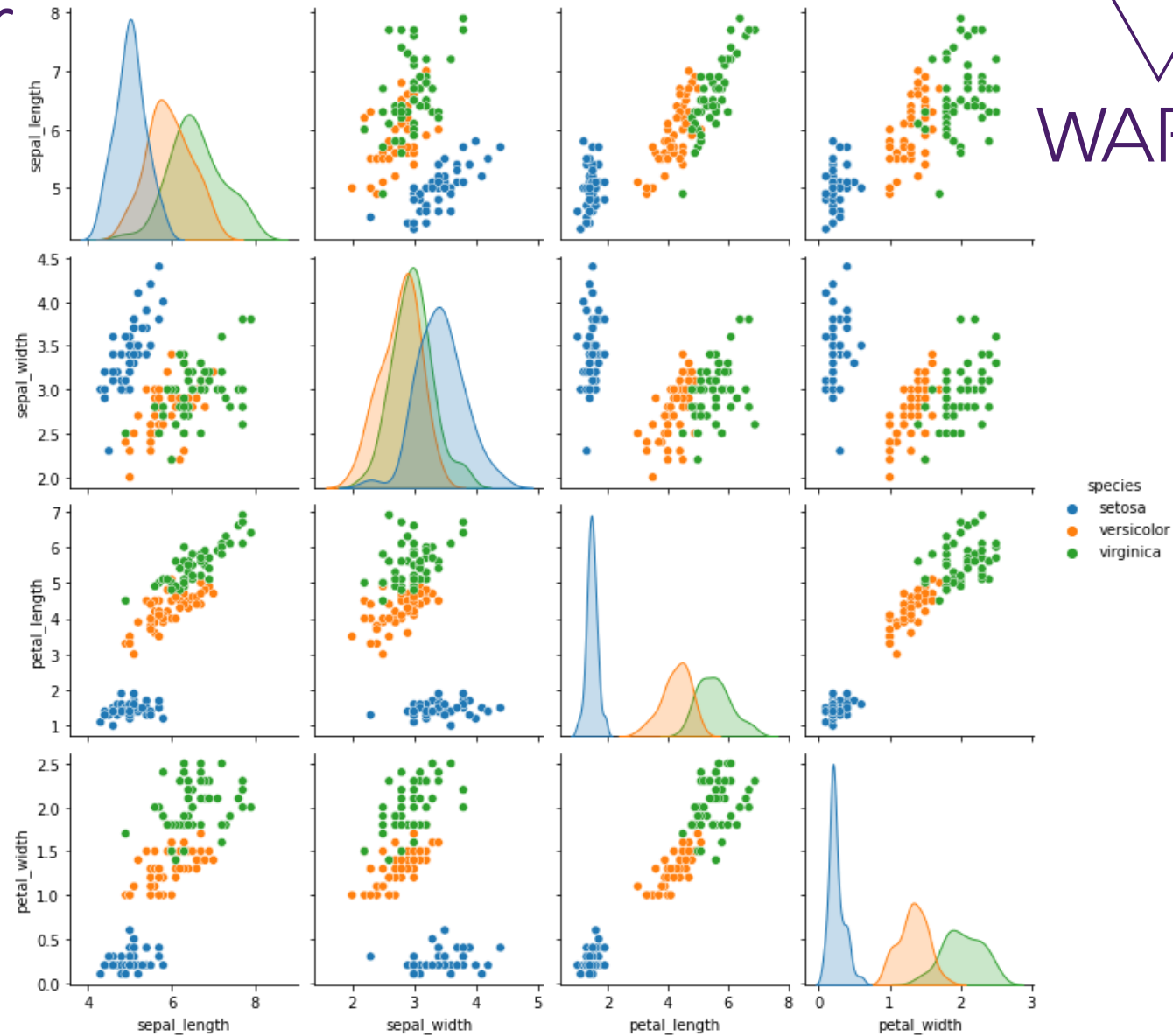


# Joint Plot



# Scatter Matrix or Pair Plot

WARWICK



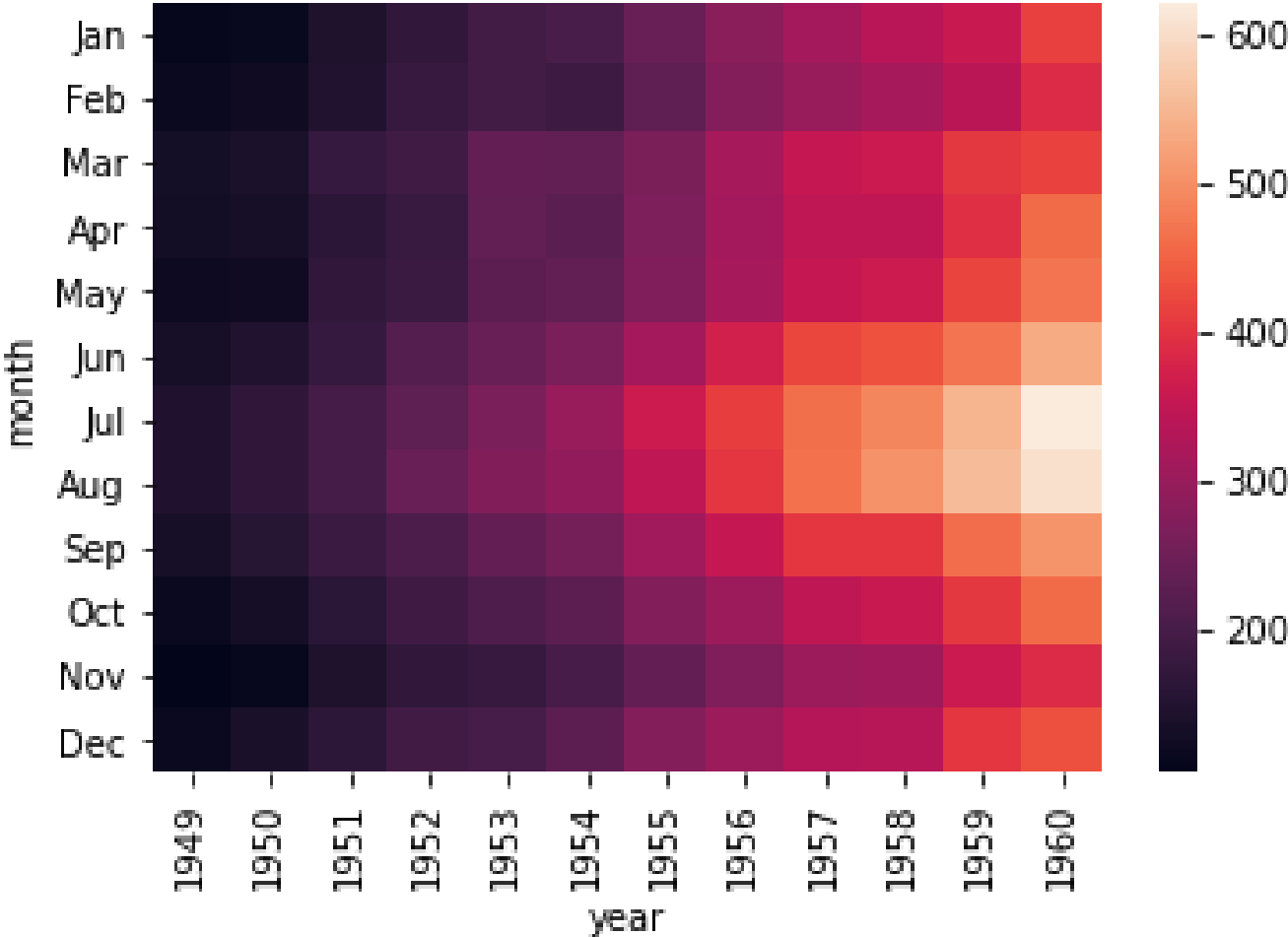
# Heat Map



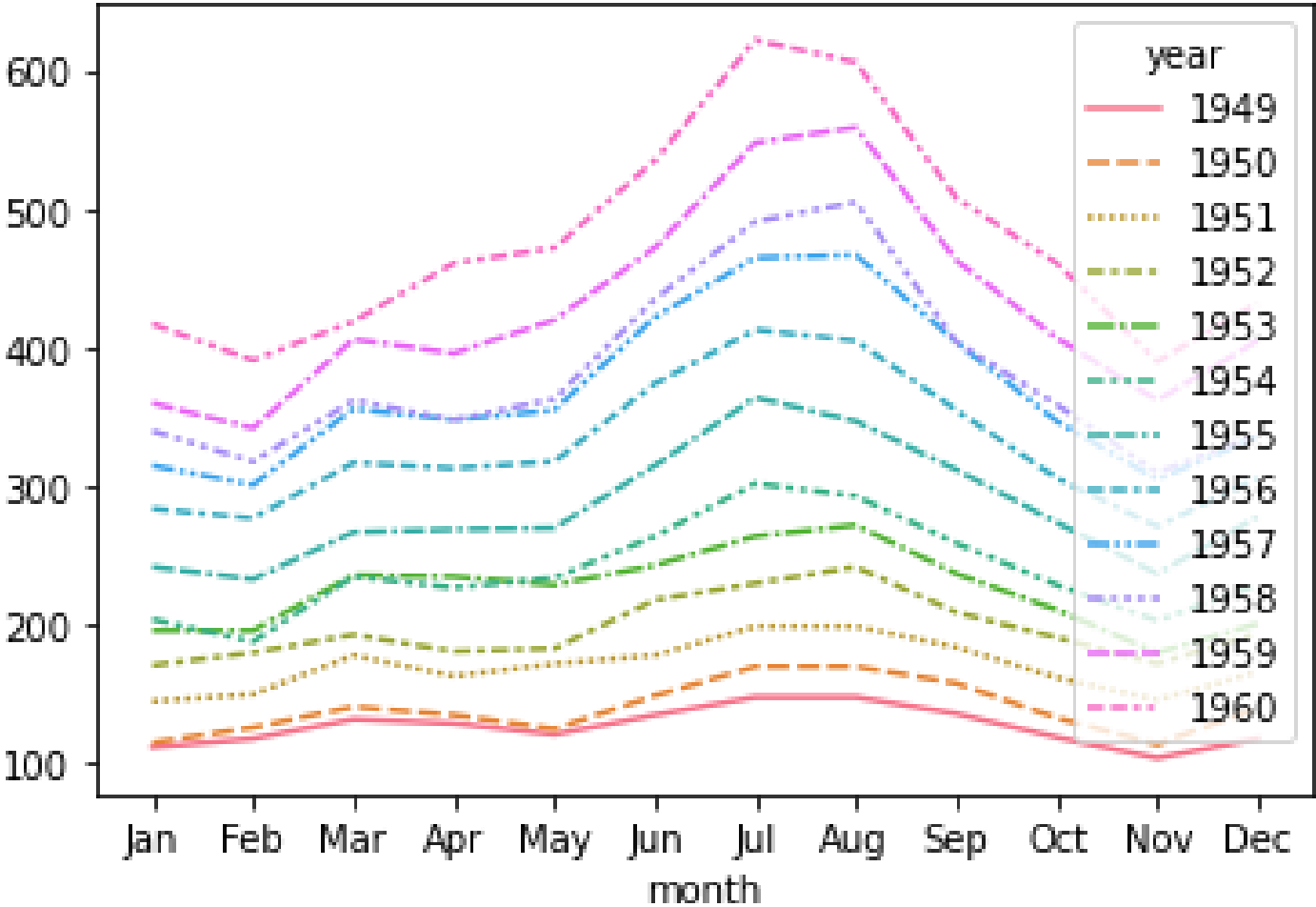
WARWICK

year	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
month												
Jan	112	115	145	171	196	204	242	284	315	340	360	417
Feb	118	126	150	180	196	188	233	277	301	318	342	391
Mar	132	141	178	193	236	235	267	317	356	362	406	419
Apr	129	135	163	181	235	227	269	313	348	348	396	461
May	121	125	172	183	229	234	270	318	355	363	420	472
Jun	135	149	178	218	243	264	315	374	422	435	472	535
Jul	148	170	199	230	264	302	364	413	465	491	548	622
Aug	148	170	199	242	272	293	347	405	467	505	559	606
Sep	136	158	184	209	237	259	312	355	404	404	463	508
Oct	119	133	162	191	211	229	274	306	347	359	407	461
Nov	104	114	146	172	180	203	237	271	305	310	362	390
Dec	118	140	166	194	201	229	278	306	336	337	405	432

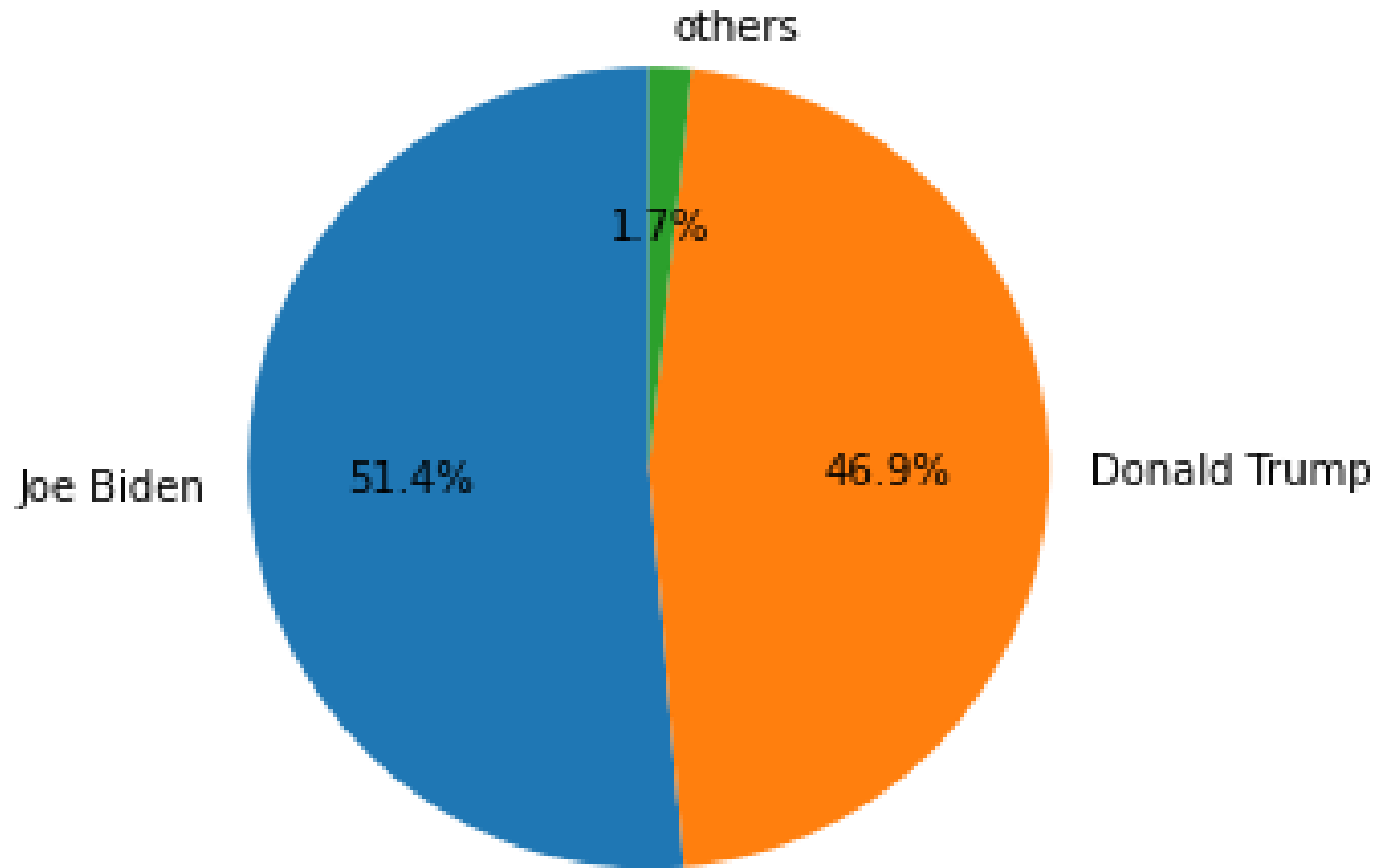
# Heat Map



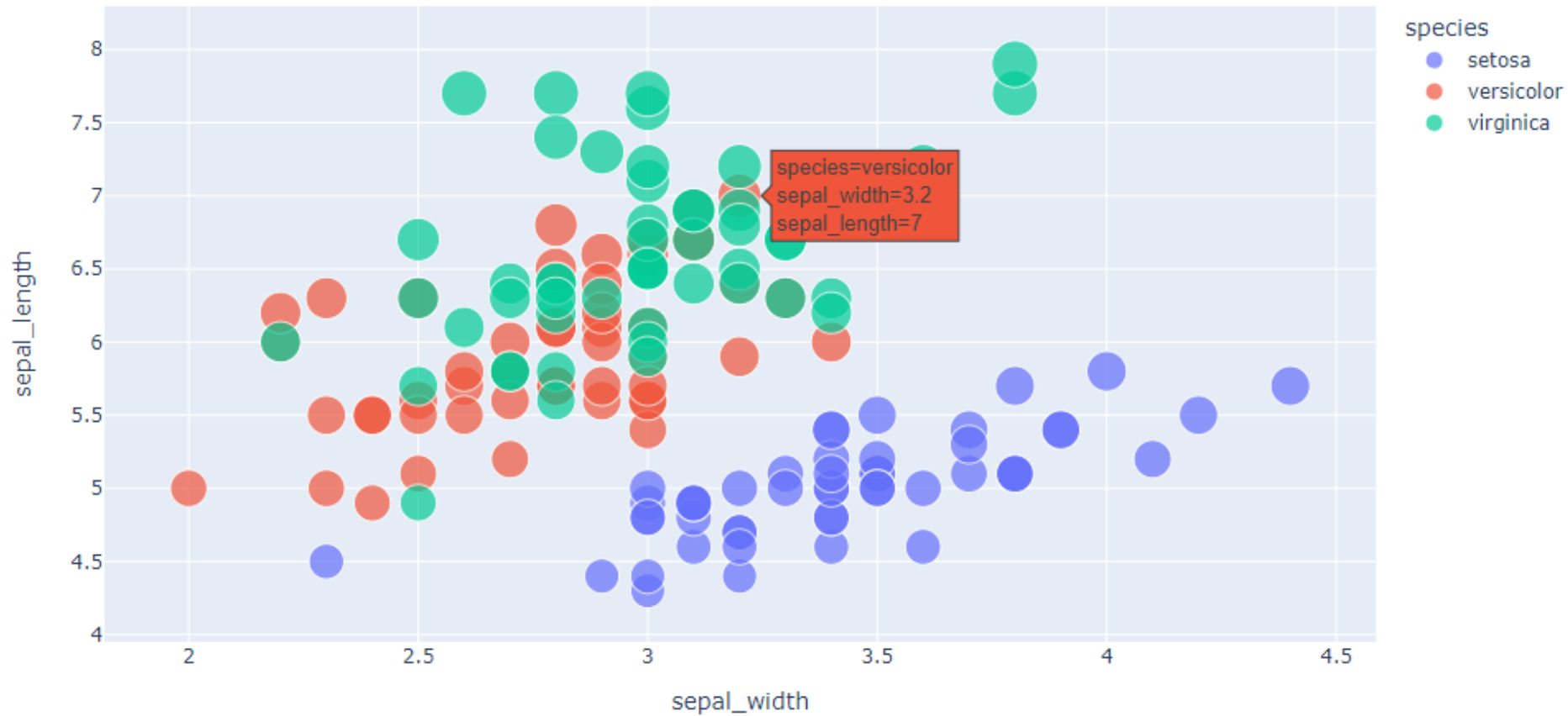
# Line Plot



# Pie Chart



# Interactive plots







“Visualization gives you answers to questions you didn’t know you had.”

– Ben Schneiderman