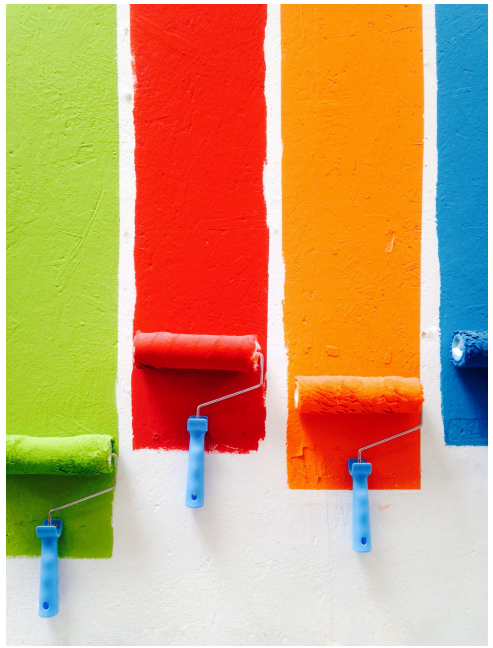
R:
A tool for big data visualization

R- Ladies talk Felicity Yi Xue. The University of Auckland

What R doing for data ?

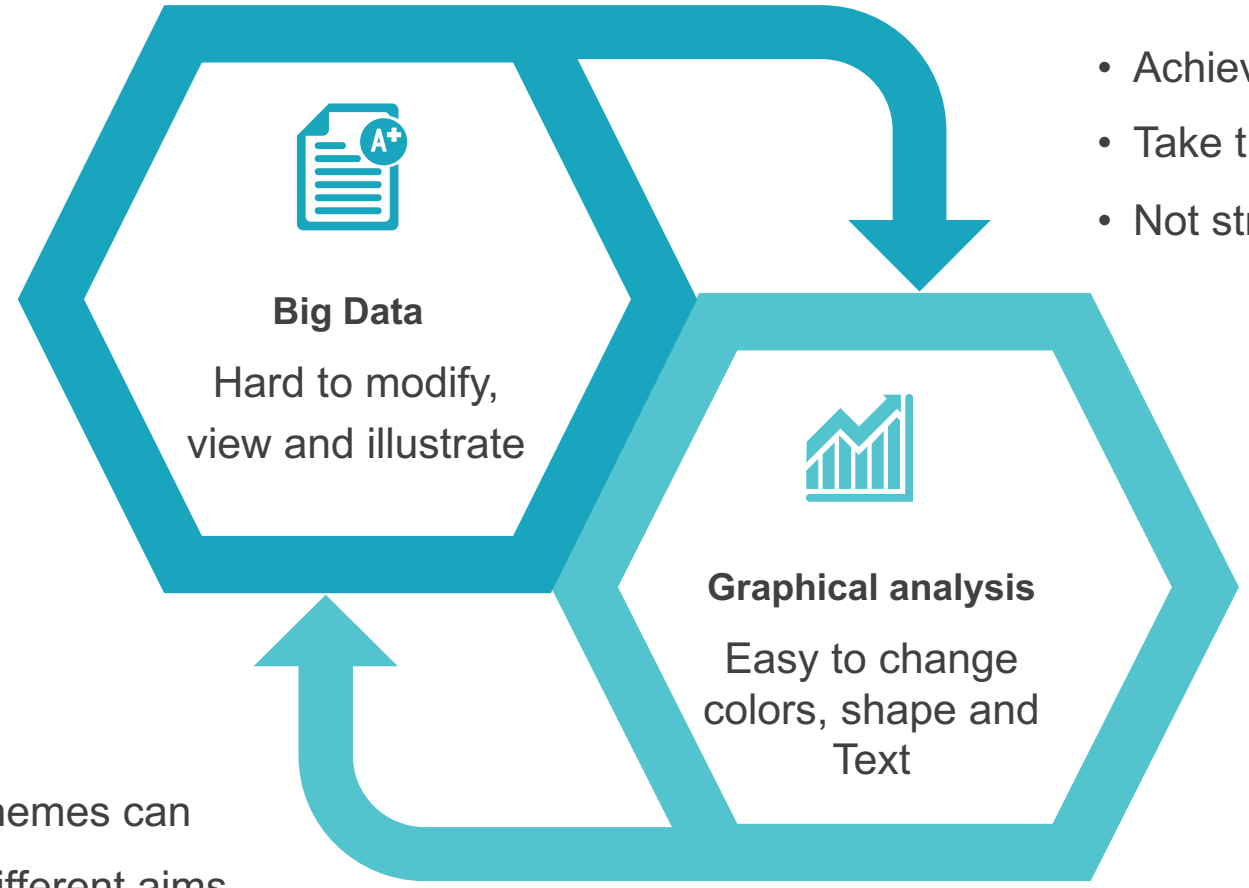


TableName	obs_daily			
lsd	station_no	rain_24h	max_air_temp	min_air_temp
2016-01-01 13:00:00	3925	1.4	30.3	11
2016-01-02 13:00:00	3925	27.2	18.7	16.5
2016-01-03 13:00:00	3925	1	22.8	16.3
2016-01-04 13:00:00	3925	0	24	6.7
2016-01-05 13:00:00	3925	0	25.2	7.8
2016-01-06 13:00:00	3925	0	26.1	9.8
2016-01-07 13:00:00	3925	28	22.4	9.4
2016-01-08 13:00:00	3925	18.8	20.9	12.7
2016-01-09 13:00:00	3925	0	21.4	11.3
2016-01-10 13:00:00	3925	0.4	20.8	11.6
2016-01-11 13:00:00	3925	0	23.2	8.4
2016-01-12 13:00:00	3925	0	22.2	12.9
2016-01-13 13:00:00	3925	0	24	11.4
2016-01-14 13:00:00	3925	0	24.6	14.2

Why?

Raw data

- Achieve from an open resource
- Take time to tidy up and customize
- Not straightforward for analyzing



Graphs and plots

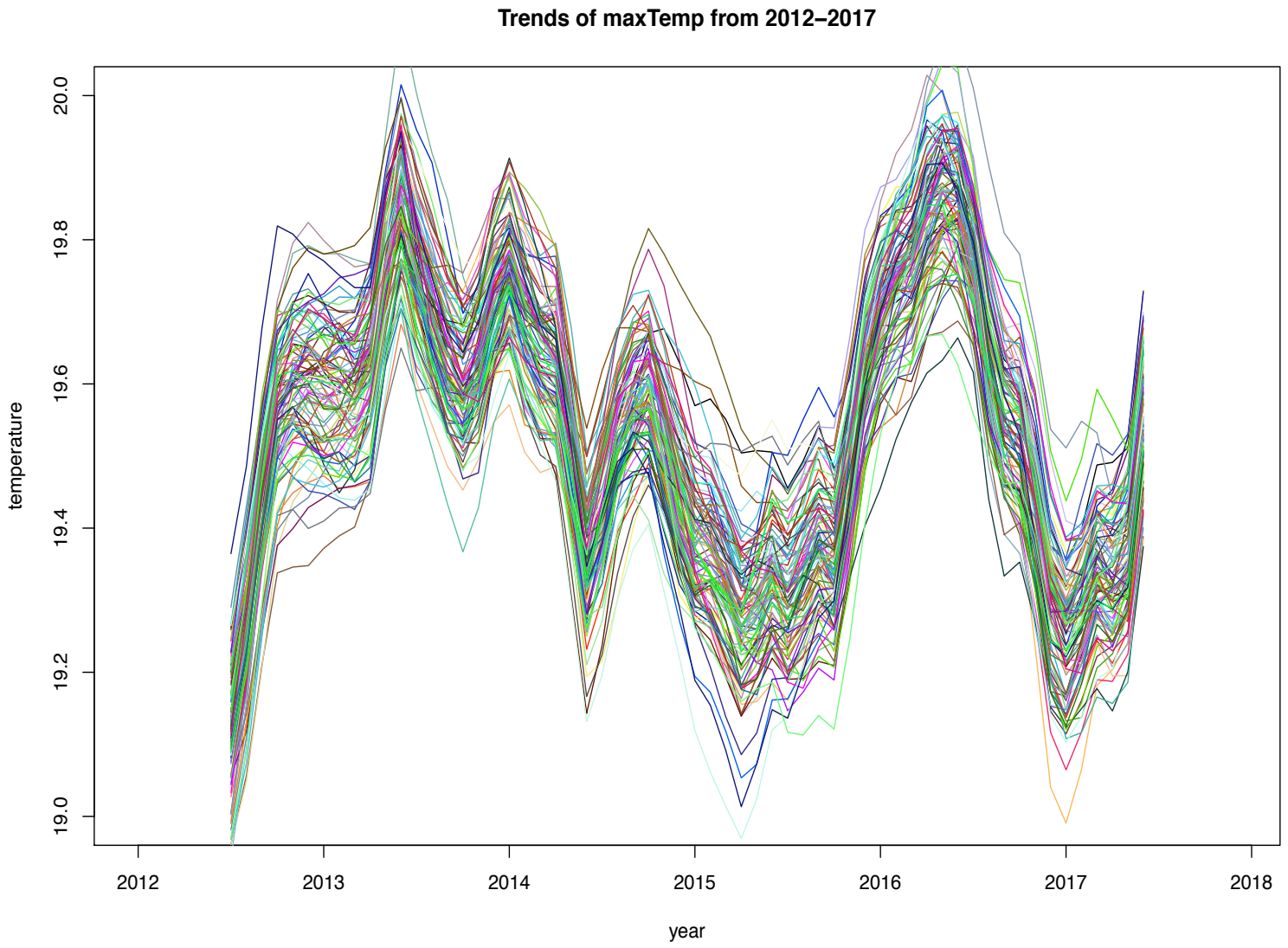
- Various graphic schemes can choose based on different aims
- Interpret data easily by visualizing the patterns

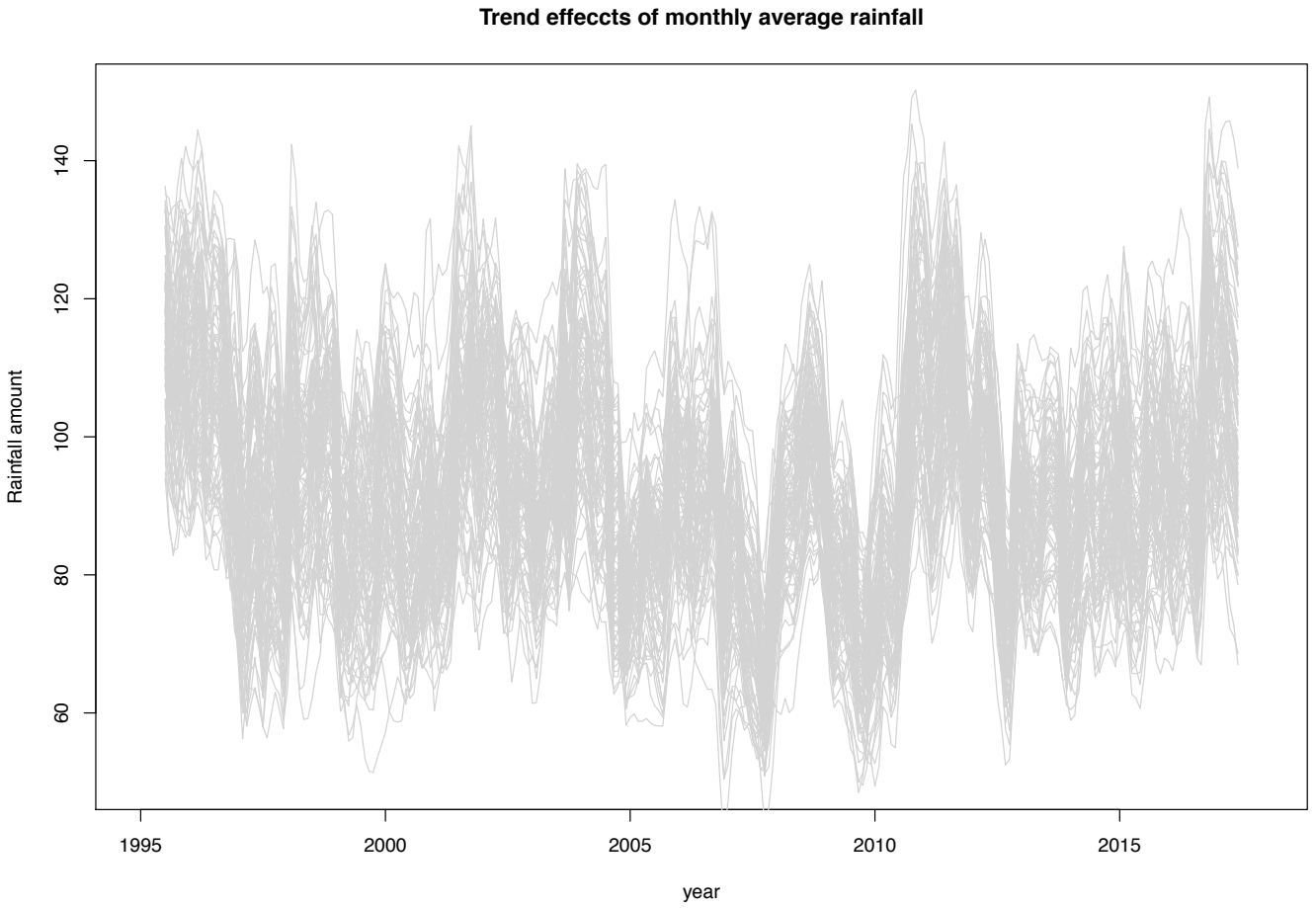
Random Colour

Rather than use default setting !

In R graphic function, for example, plot(). The default black lines are not friendly for distinguishing from each other.

- Here we interested in the comparison of annually tendency





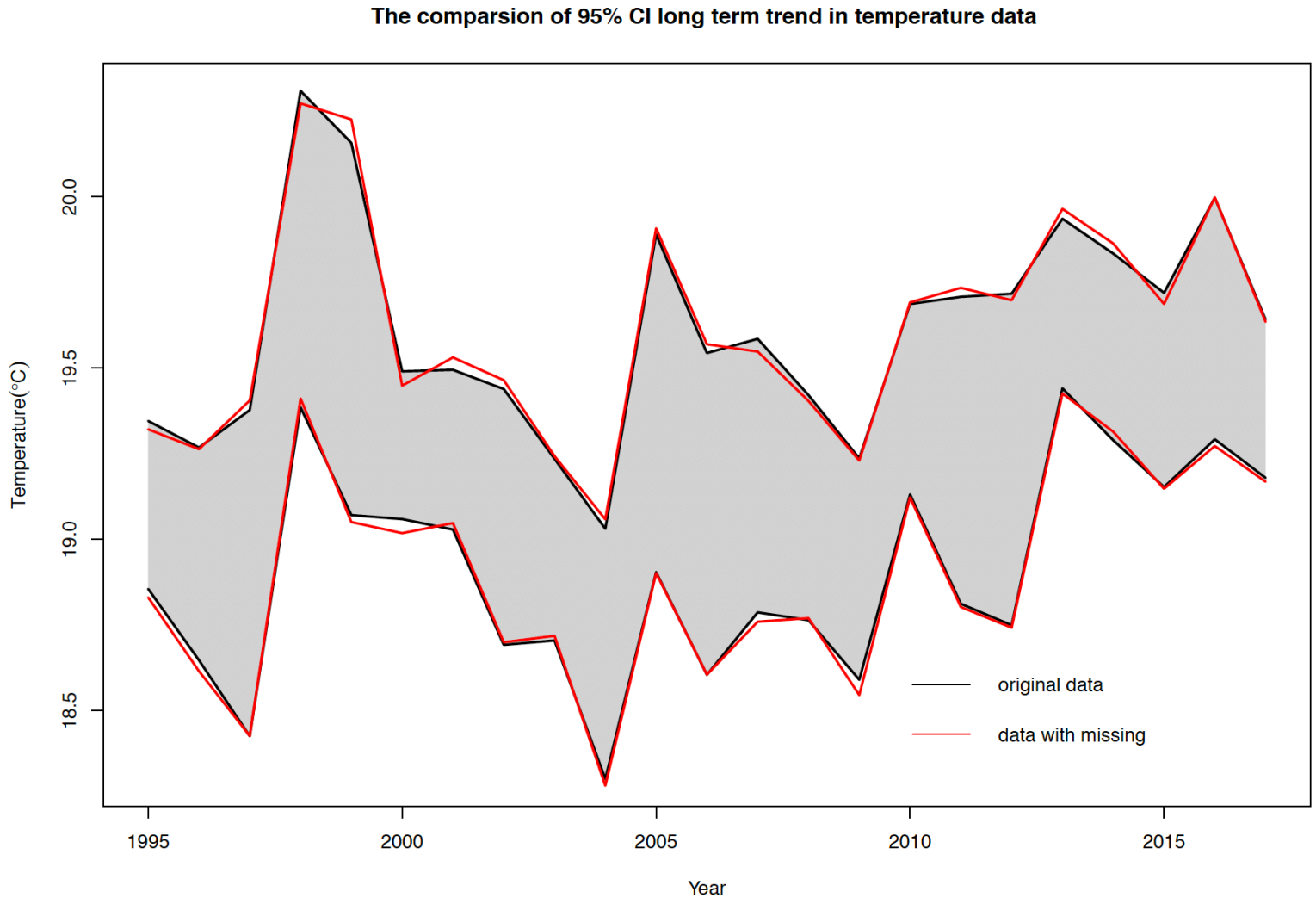
Then we gather all the sets of annual data and shaded them in grey.

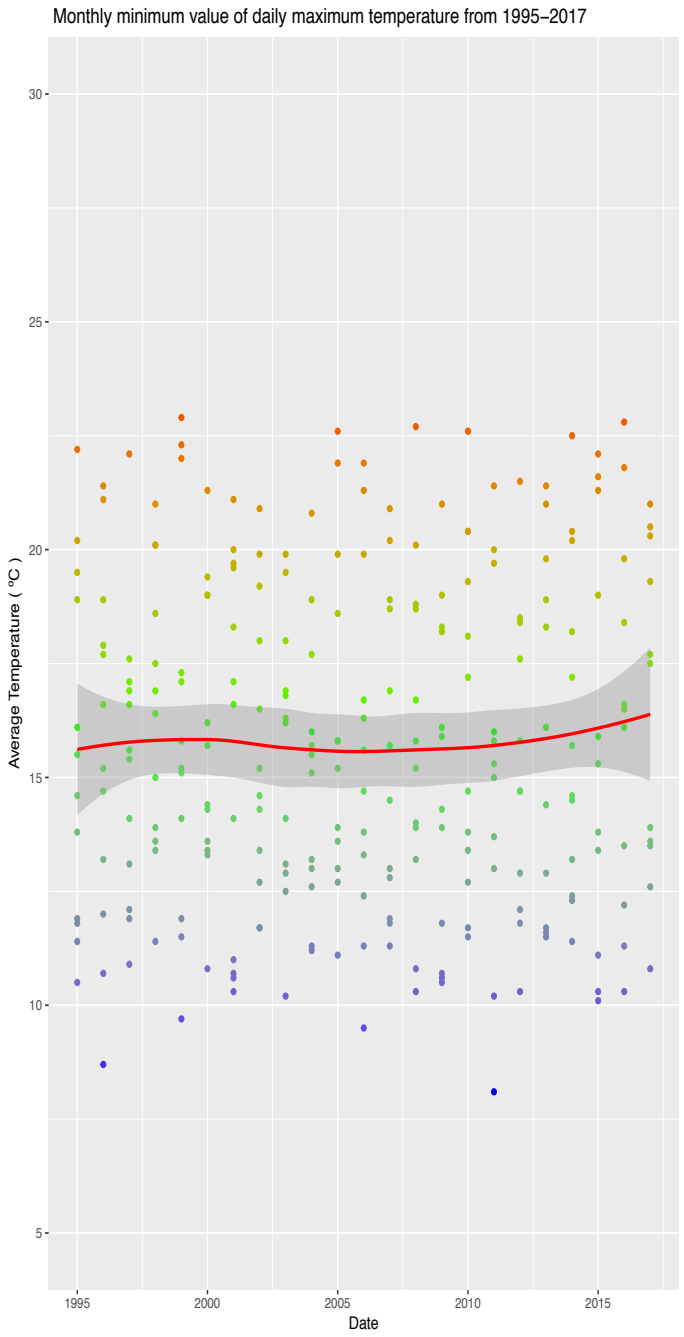
- To weaken the pattern of individual set.
- To recognize the whole picture of average pattern.

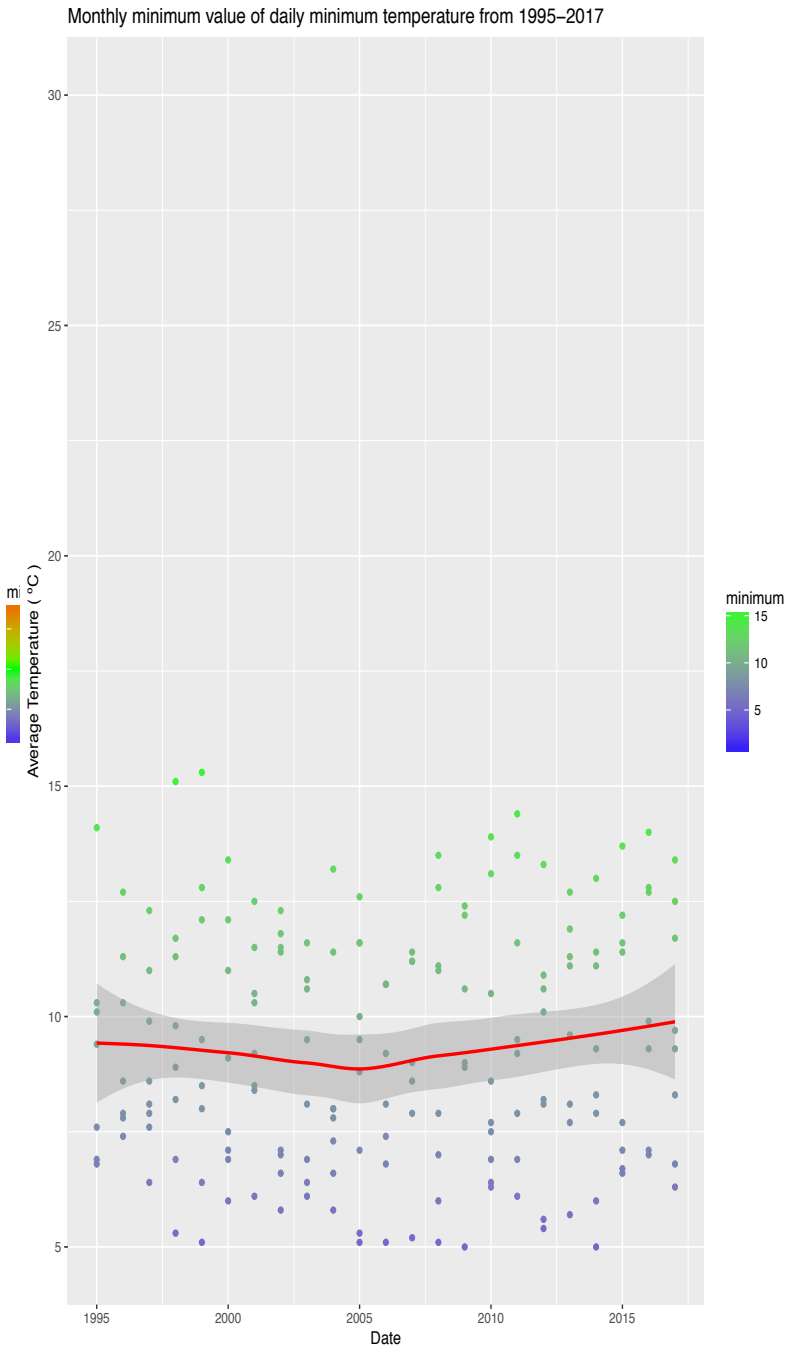
Data Visualize

Compare the difference in one graph

- ✓Analysis of 95% confidence interval for time series data.
- ✓The area of 95% CI for original data are shaded in grey
- ✓The red line shows the shape of artificial data that we used to compare to the original data







- Also, analyzing the data by using simple function to produce other aspects of understanding the information.

- Here, we pay attention on the extreme values over a data set, like maxima and minima.


THANK YOU
Questions?