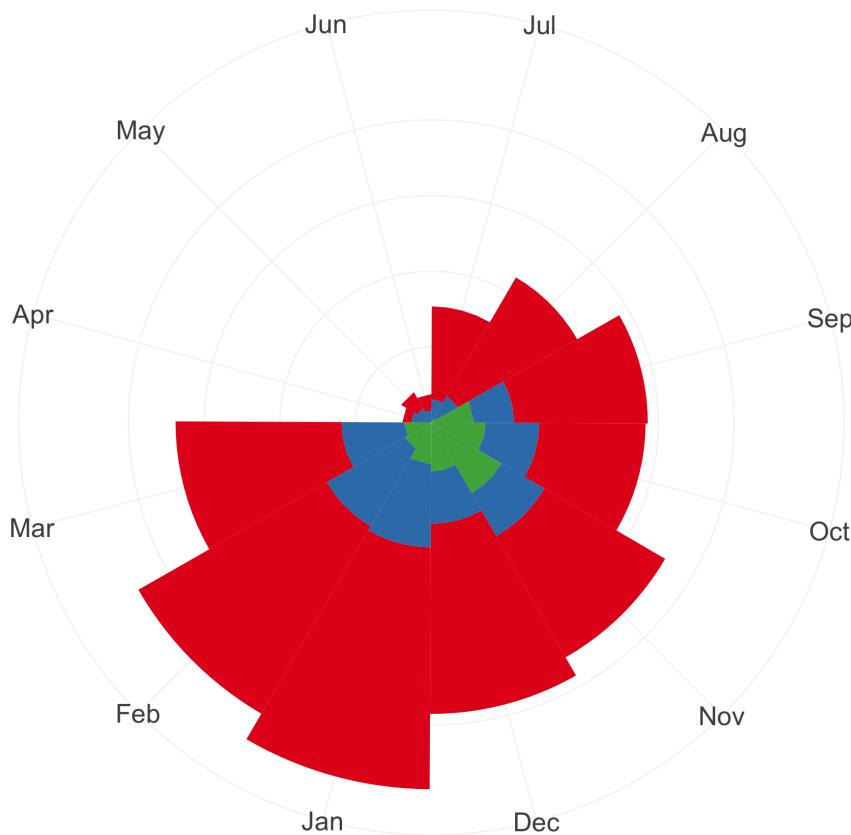


Data Visualization

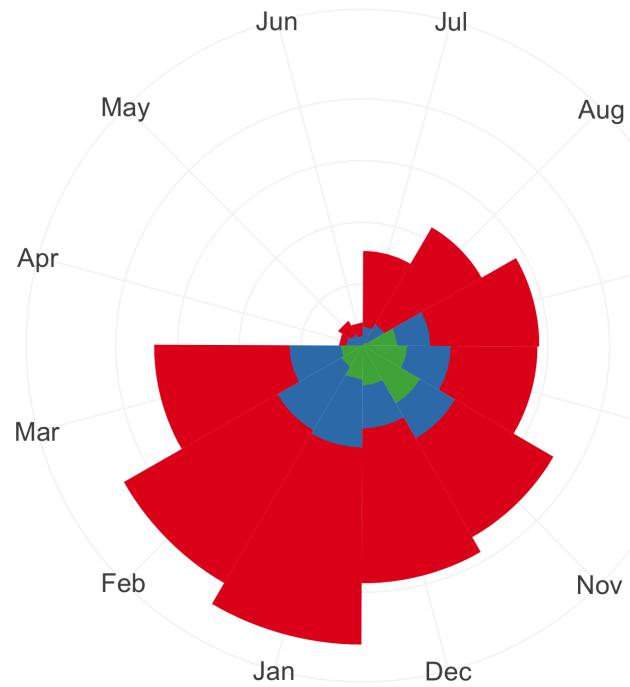
CORH 205

Mathew Vis-Dunbar

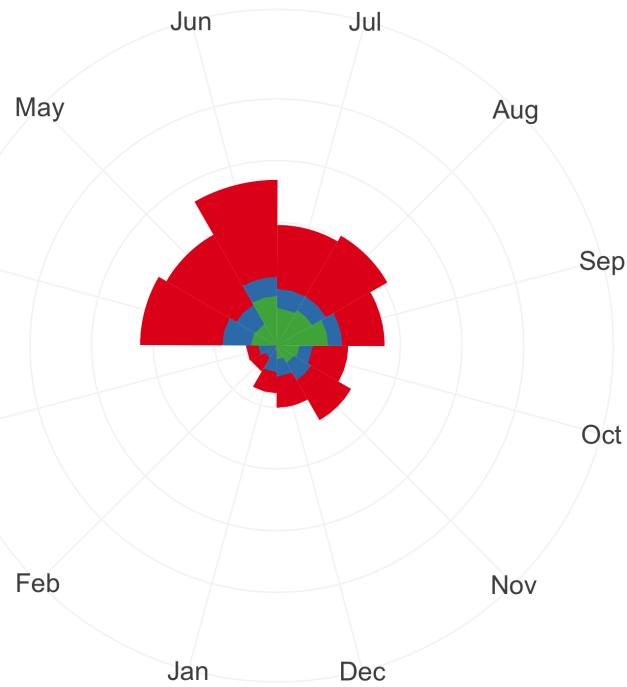
November 2023



April 1854 to March 1855

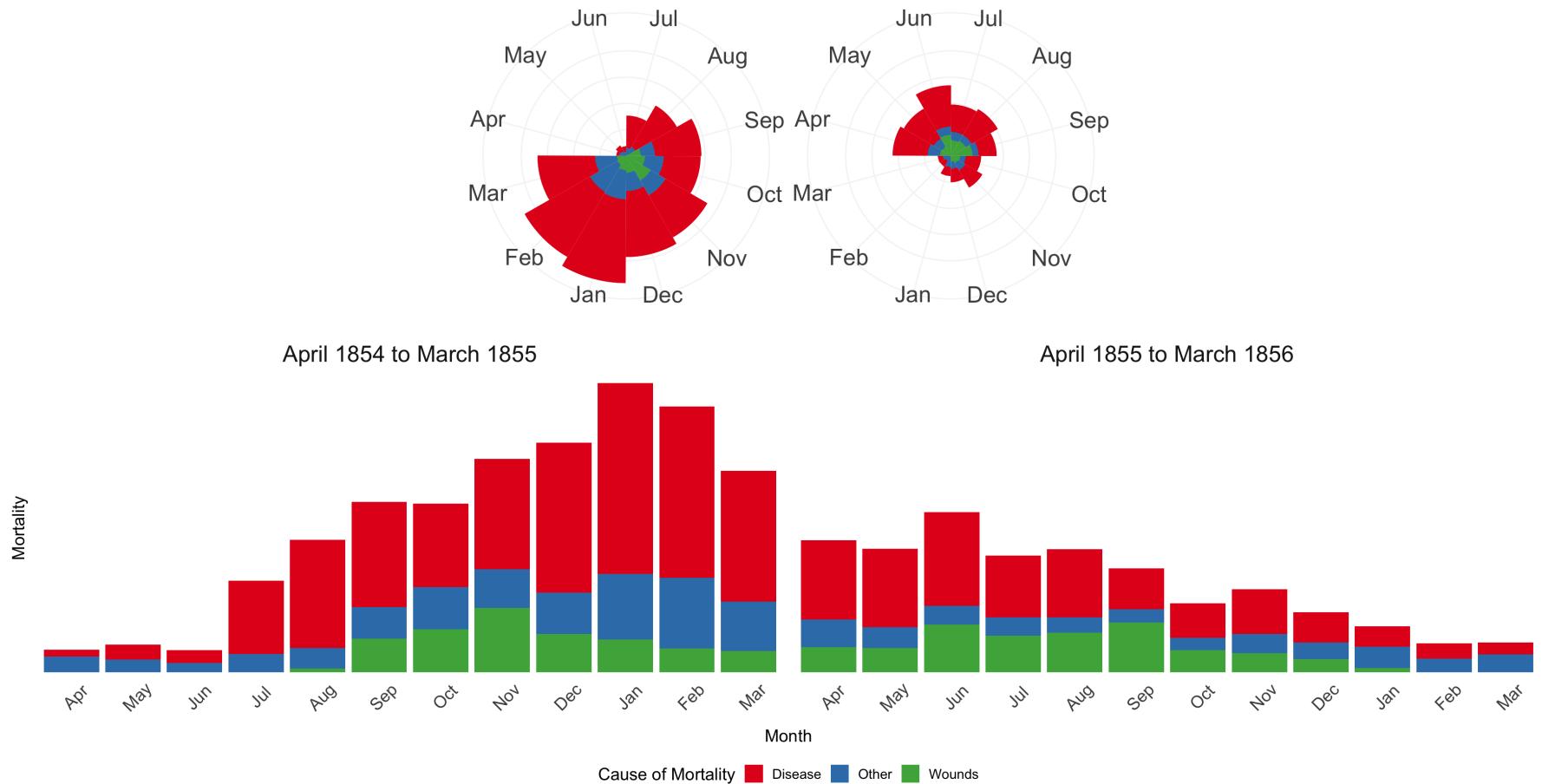


April 1855 to March 1856

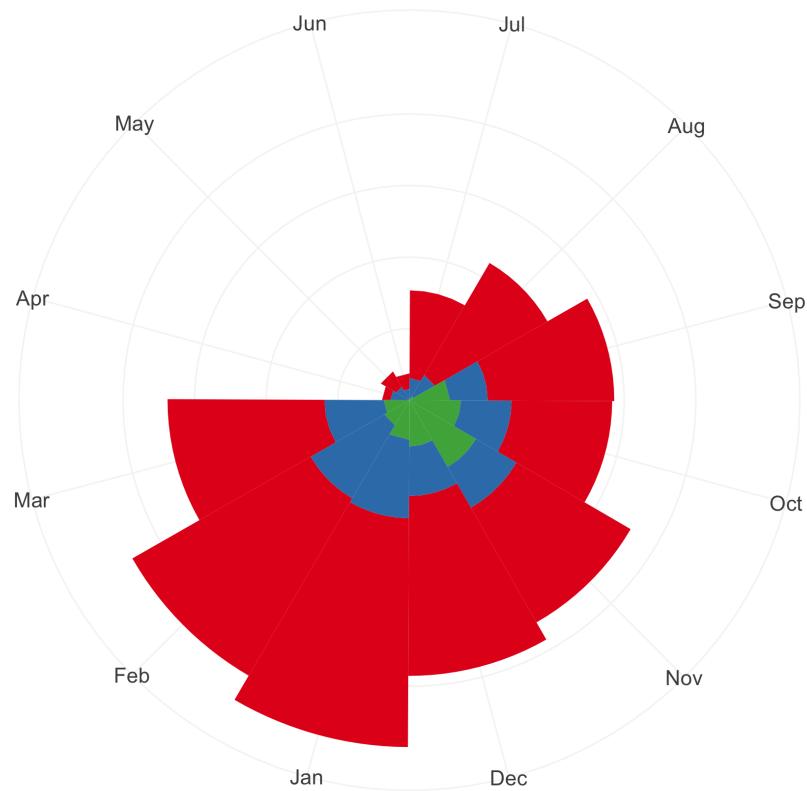


Cause of Mortality

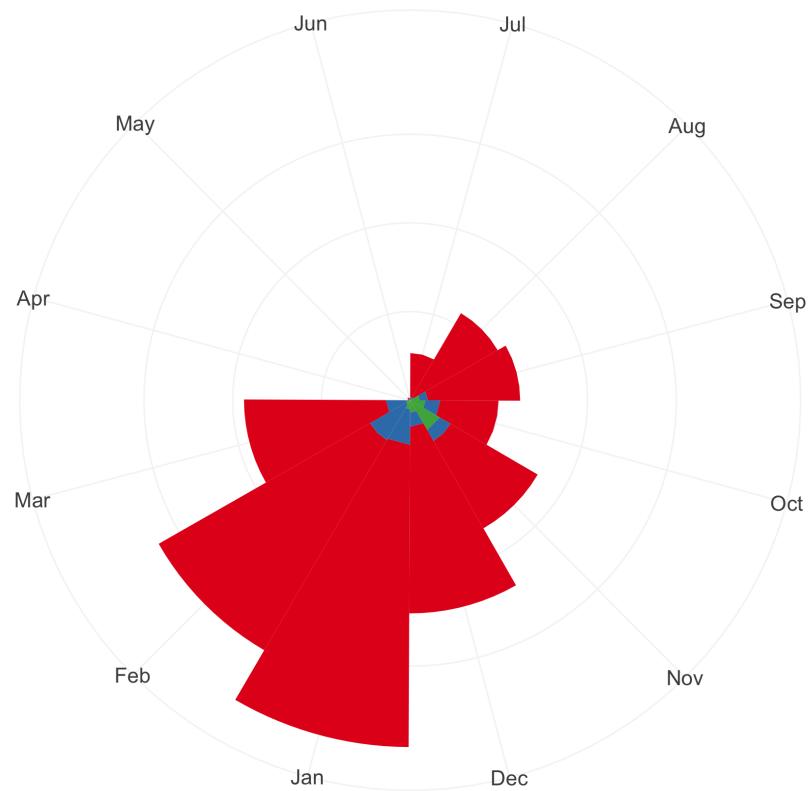
■ Disease ■ Other ■ Wounds

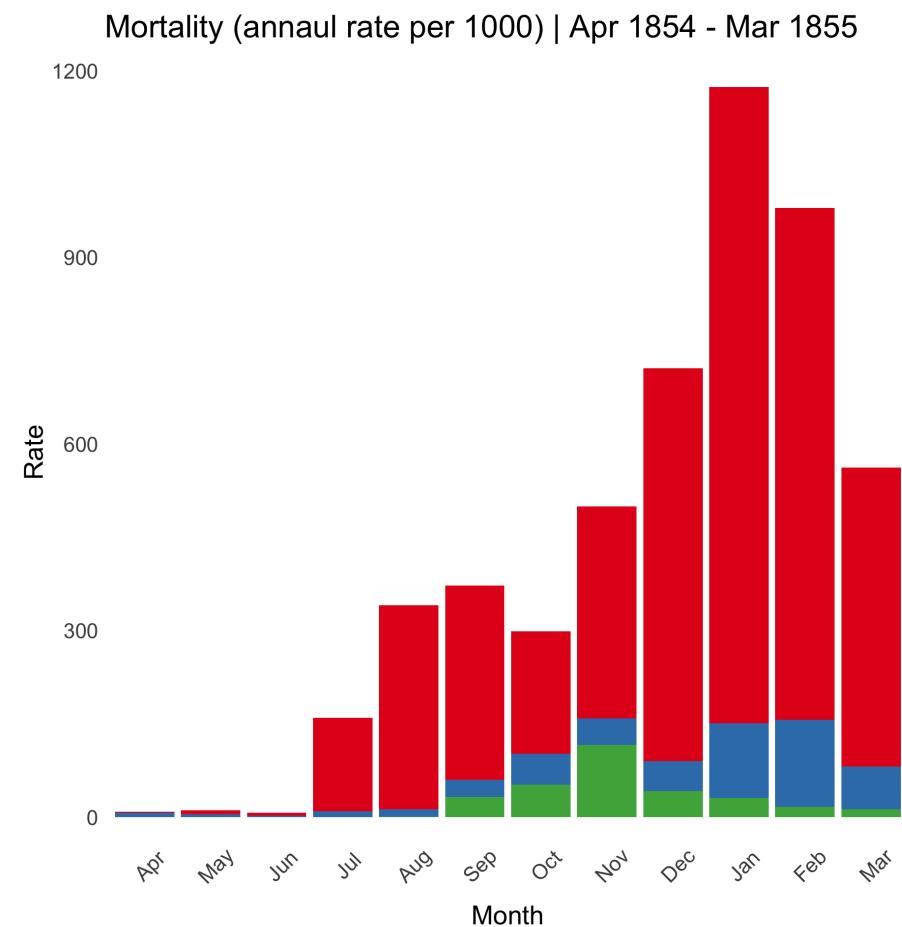
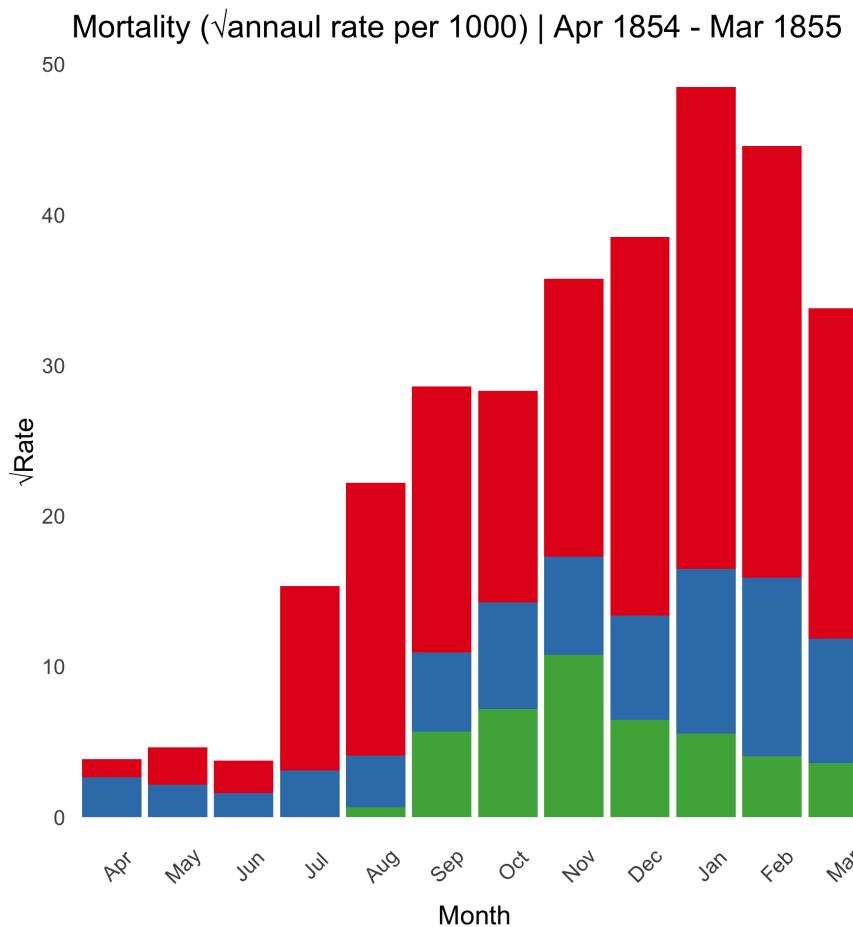


Mortality (\sqrt{annaul} rate per 1000) | Apr 1854 - Mar 1855



Mortality (annaul rate per 1000) | Apr 1854 - Mar 1855





Themes

- A little bit of theory
- Why Visualize Data
- Communicating Data vs a Message
- Data Types
- Application of the theory
- Common Visualizations
- Layering Variables
- Data Transformations

first ... what is data visualization?

Data visualization turns data into visual information. We could turn data into auditory or tactile information as well.

Data visualization turns data into visual information. We could turn data into auditory or tactile information as well.

This involves abstraction – the data and patterns may be represented as shapes (form), categorized with colours, and connected via position.

Preattentive Attributes

Preattentive Attributes

Even though an object as a whole might take some conscious effort to identify, the basic visual attributes that combine to make up that object are perceived without any conscious effort.

Stephen Few (2004).[Tapping the Power of Visual Perception](#).

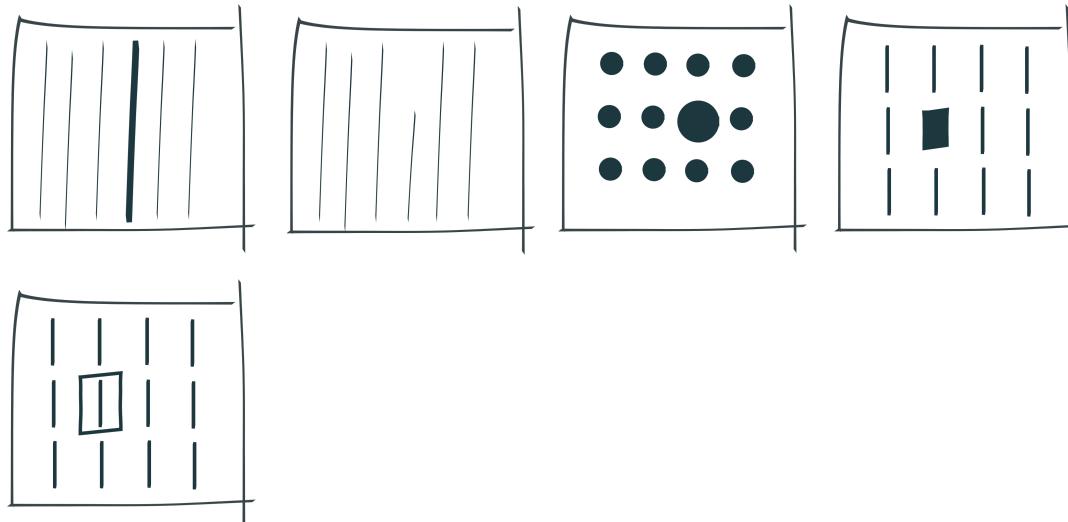
Preattentive Attributes

The things we process before we're truly cognizant of the information.

Preattentive Attributes

The things we process before we're truly cognizant of the information.

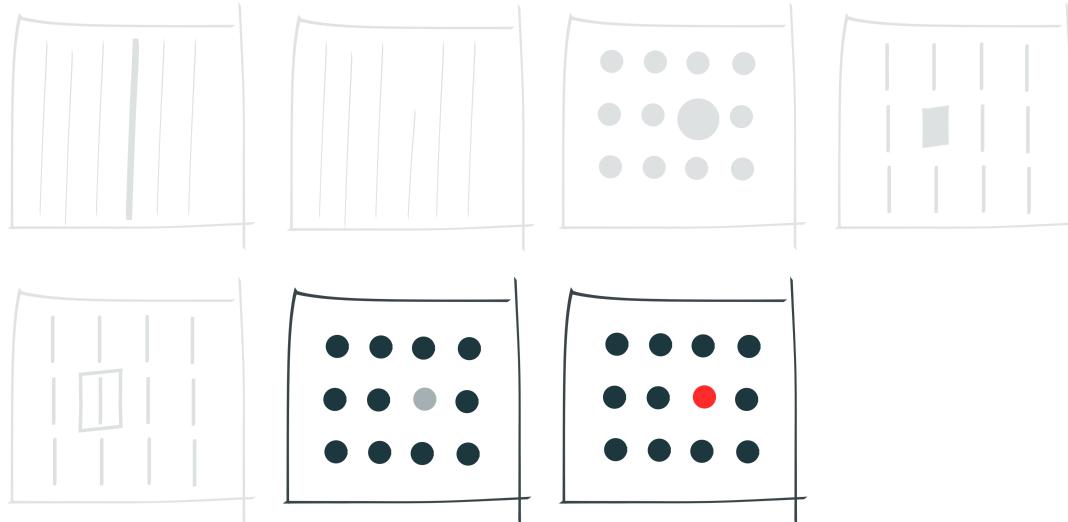
Form.



Preattentive Attributes

The things we process before we're truly cognizant of the information.

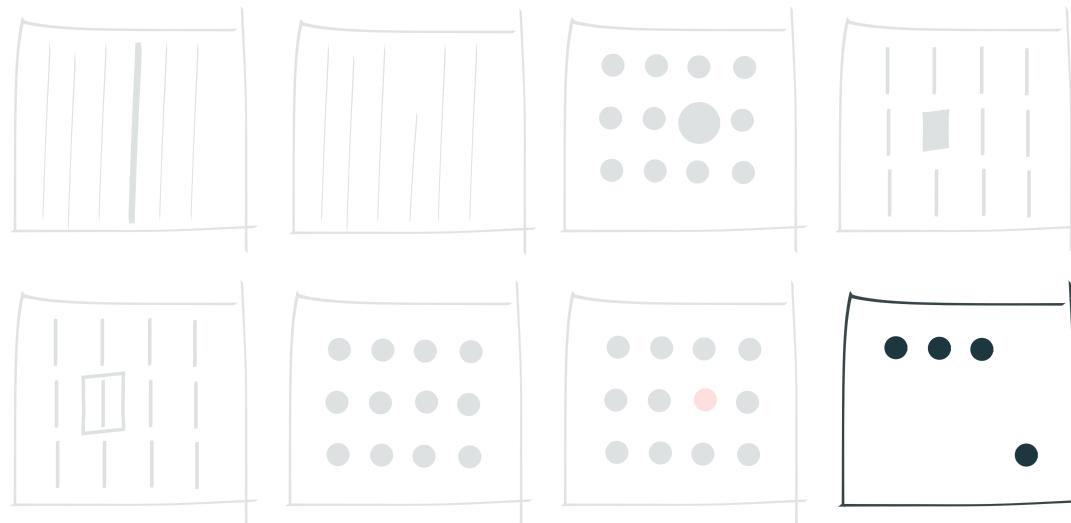
Form. Colour.



Preattentive Attributes

The things we process before we're truly cognizant of the information.

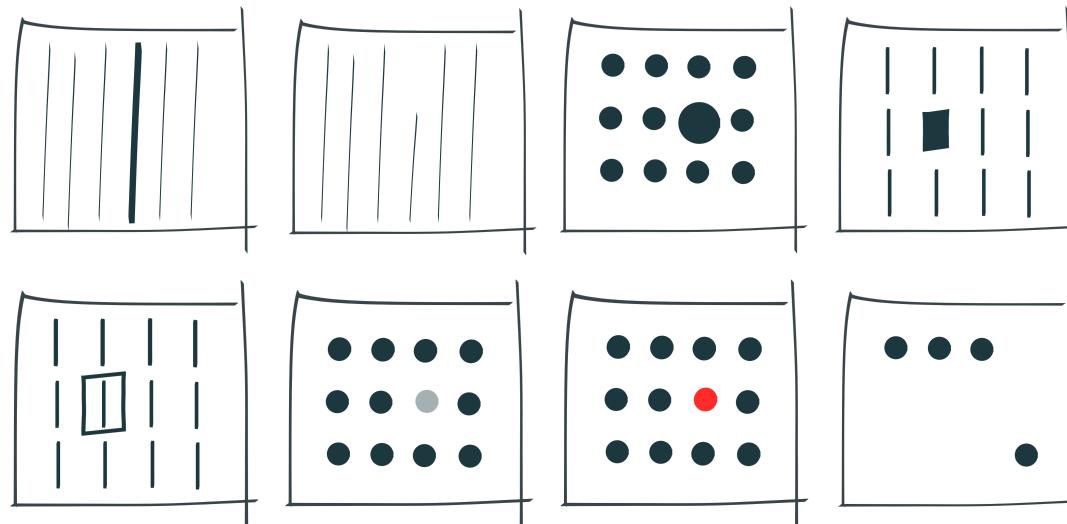
Form. Colour. Position.



Preattentive Attributes

The things we process before we're truly cognizant of the information.

Form. Colour. Position.



We can do a little test...

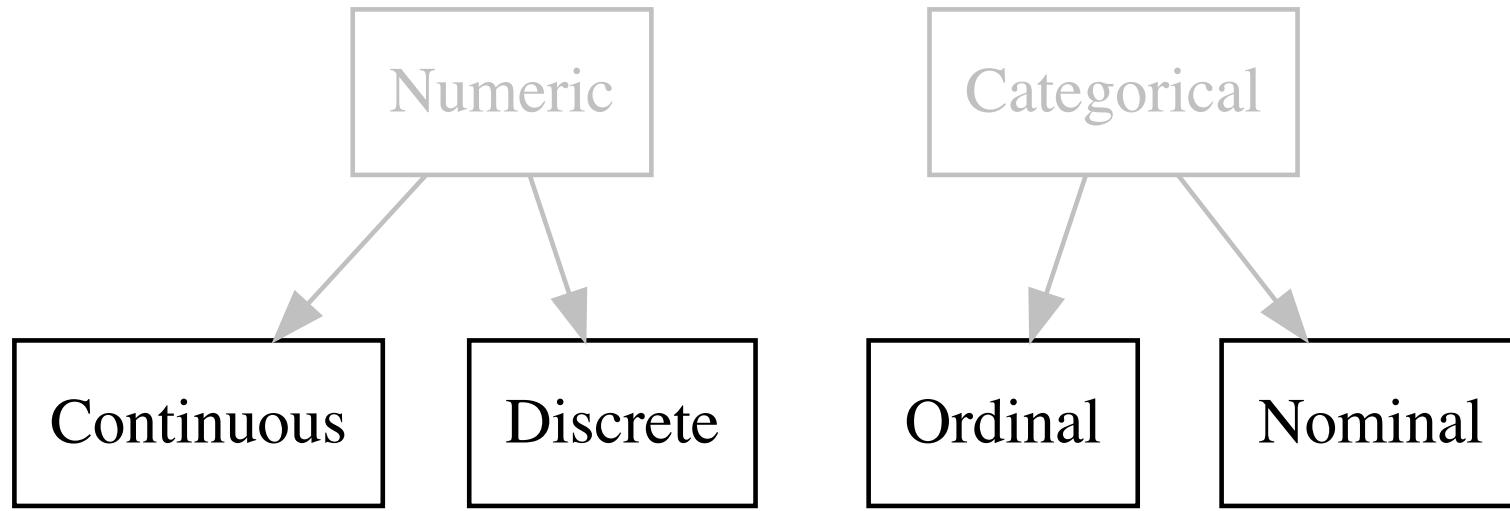
E	5	9	A	5	C	C	9	E	2	C	C
4	4	8	8	E	6	9	5	C	5	8	9
8	C	9	9	A	9	C	C	4	5	5	E
4	C	9	5	6	E	6	E	4	2	C	8
4	8	A	C	6	9	5	A	6	2	2	5
6	5	4	4	9	9	9	2	8	2	4	E
8	A	E	8	8	9	9	5	6	9	A	E
5	9	C	9	A	4	2	5	9	2	5	9
A	6	8	A								

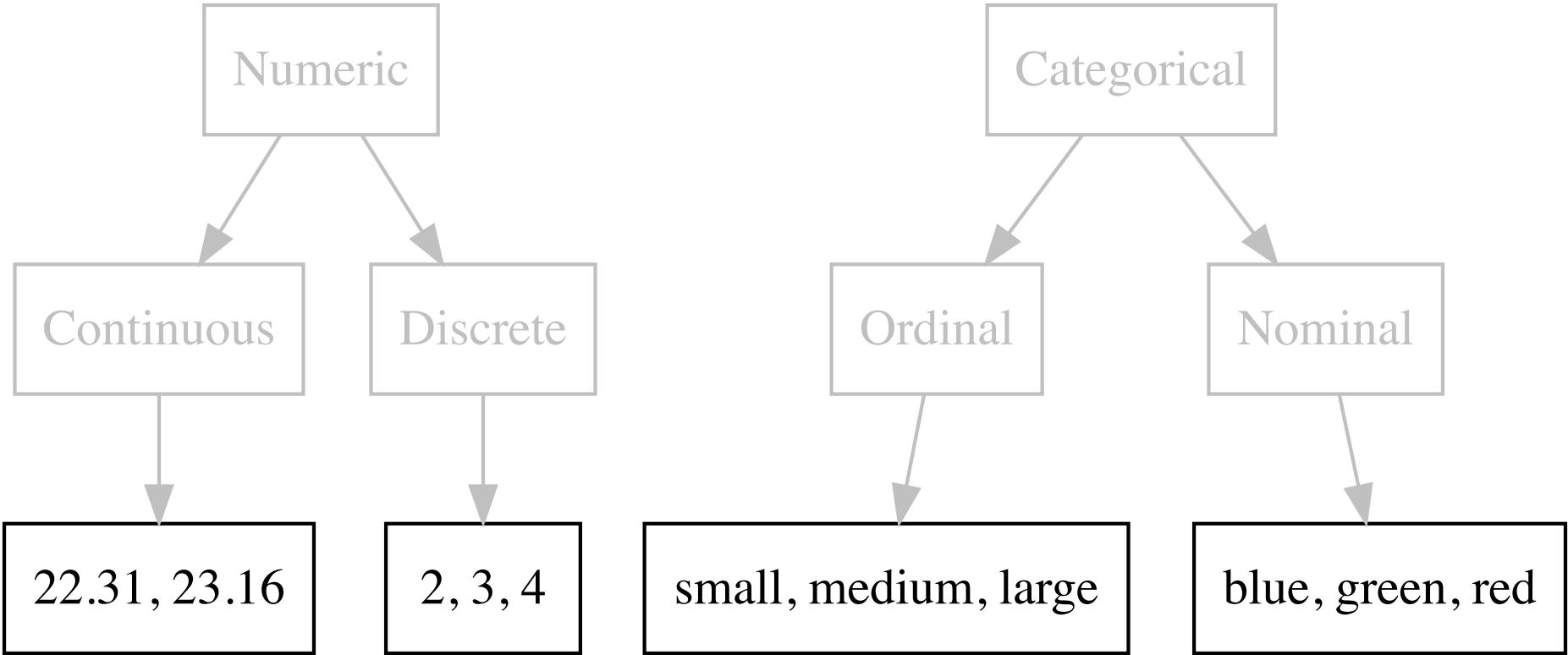
5	5	5	C	E	C	C	9	E	A	A	A
5	5	5	8	E	6	9	9	C	A	A	A
5	5	5	9	C	9	C	C	4	A	A	A
5	5	5	4	6	E	6	E	4	2	C	8
5	5	2	C	6	9	C	4	6	2	2	8
6	8	4	4	9	9	9	2	8	2	4	E
8	8	E	8	8	9	9	9	6	9	9	E
4	9	C	9	4	4	2	C	9	2	9	9
8	6	8	E								

second ... data types

Numeric

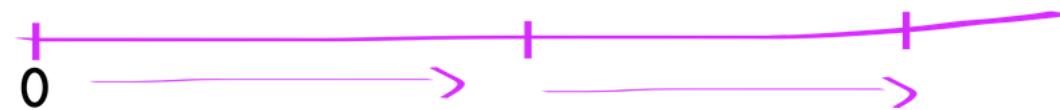
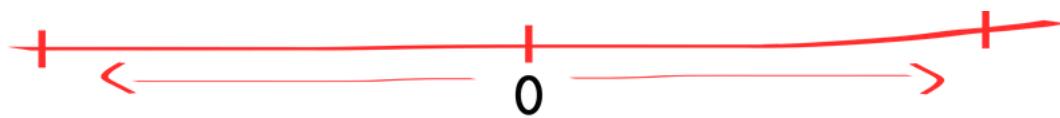
Categorical



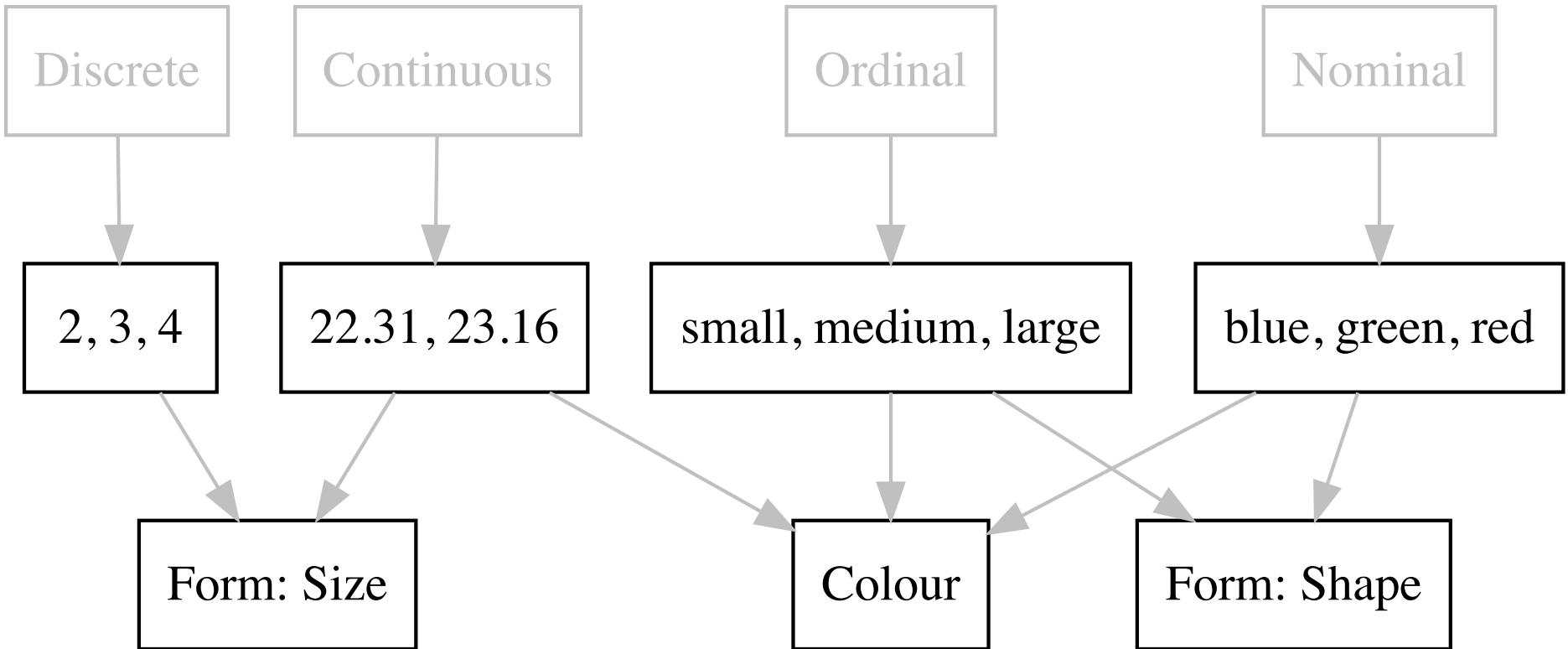


Interval

Ratio



Preattentive attributes + Data types



third ... why visualize data?

Data visualization is the graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication.

Stephen Few. [Data Visualization for Human Perception](#).

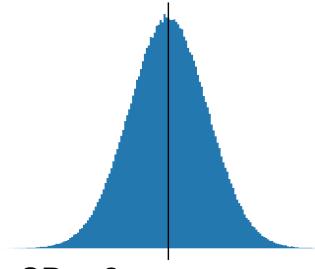
Visualization in Sense Making

Data are generally summarized using test statistics; a single number to describe an entire data set.

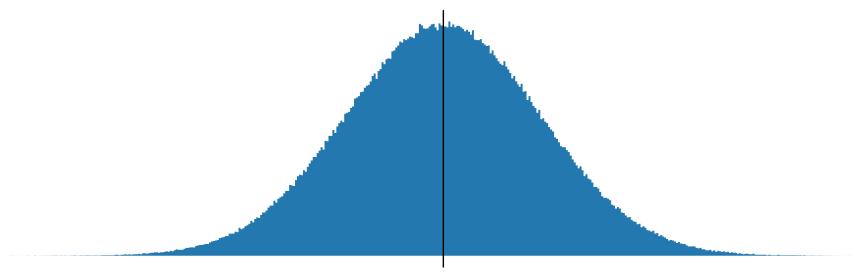
Averages means describing central tendencies

Spreads standard deviations describing the average spread of the data

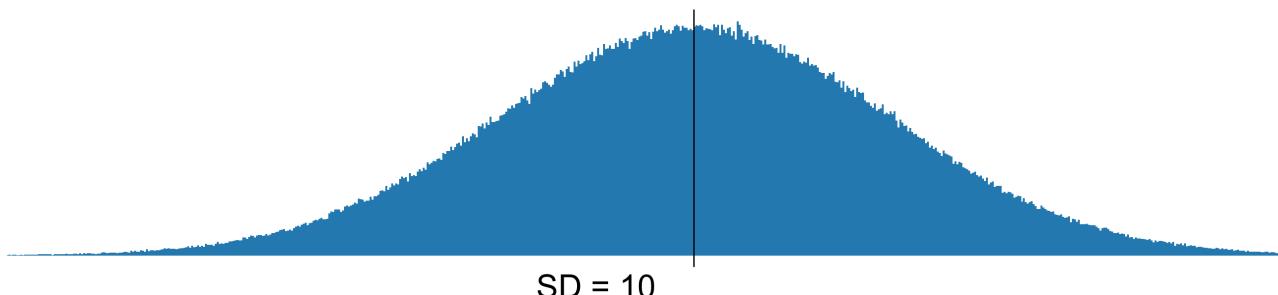
Relationships correlation coefficients describing the direction of a relationship



SD = 2



SD = 5



SD = 10

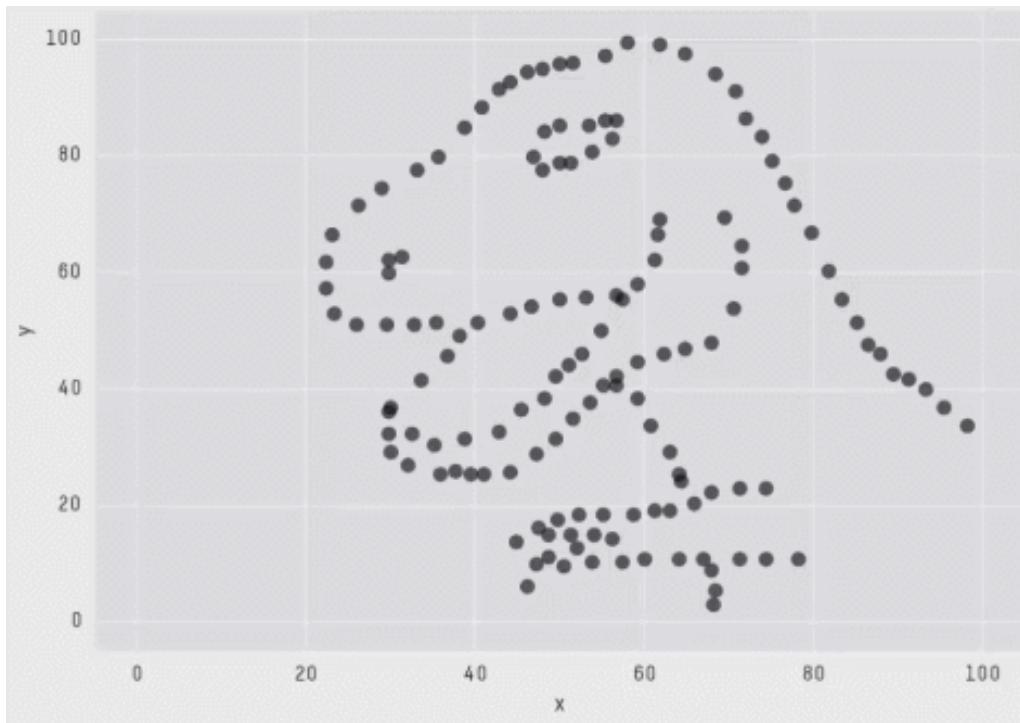
Dinosaurus Dozen

dataset	mean_x	mean_y	sd_x	sd_y	correlation
away	54.26610	47.83472	16.76983	26.93974	-0.0641284
bullseye	54.26873	47.83082	16.76924	26.93573	-0.0685864
circle	54.26732	47.83772	16.76001	26.93004	-0.0683434
dino	54.26327	47.83225	16.76514	26.93540	-0.0644719
dots	54.26030	47.83983	16.76774	26.93019	-0.0603414
h_lines	54.26144	47.83025	16.76590	26.93988	-0.0617148
high_lines	54.26881	47.83545	16.76670	26.94000	-0.0685042
slant_down	54.26785	47.83590	16.76676	26.93610	-0.0689797
slant_up	54.26588	47.83150	16.76885	26.93861	-0.0686092
star	54.26734	47.83955	16.76896	26.93027	-0.0629611
v_lines	54.26993	47.83699	16.76996	26.93768	-0.0694456
wide_lines	54.26692	47.83160	16.77000	26.93790	-0.0665752
x_shape	54.26015	47.83972	16.76996	26.93000	-0.0655833

Anscombe, F.J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1), 17. <https://doi.org/10.2307/2682899>

Matejka, J., & Fitzmaurice, G. (2017). Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1290–1294.

<https://doi.org/10.1145/3025453.3025912>



X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

Visualization for Communication

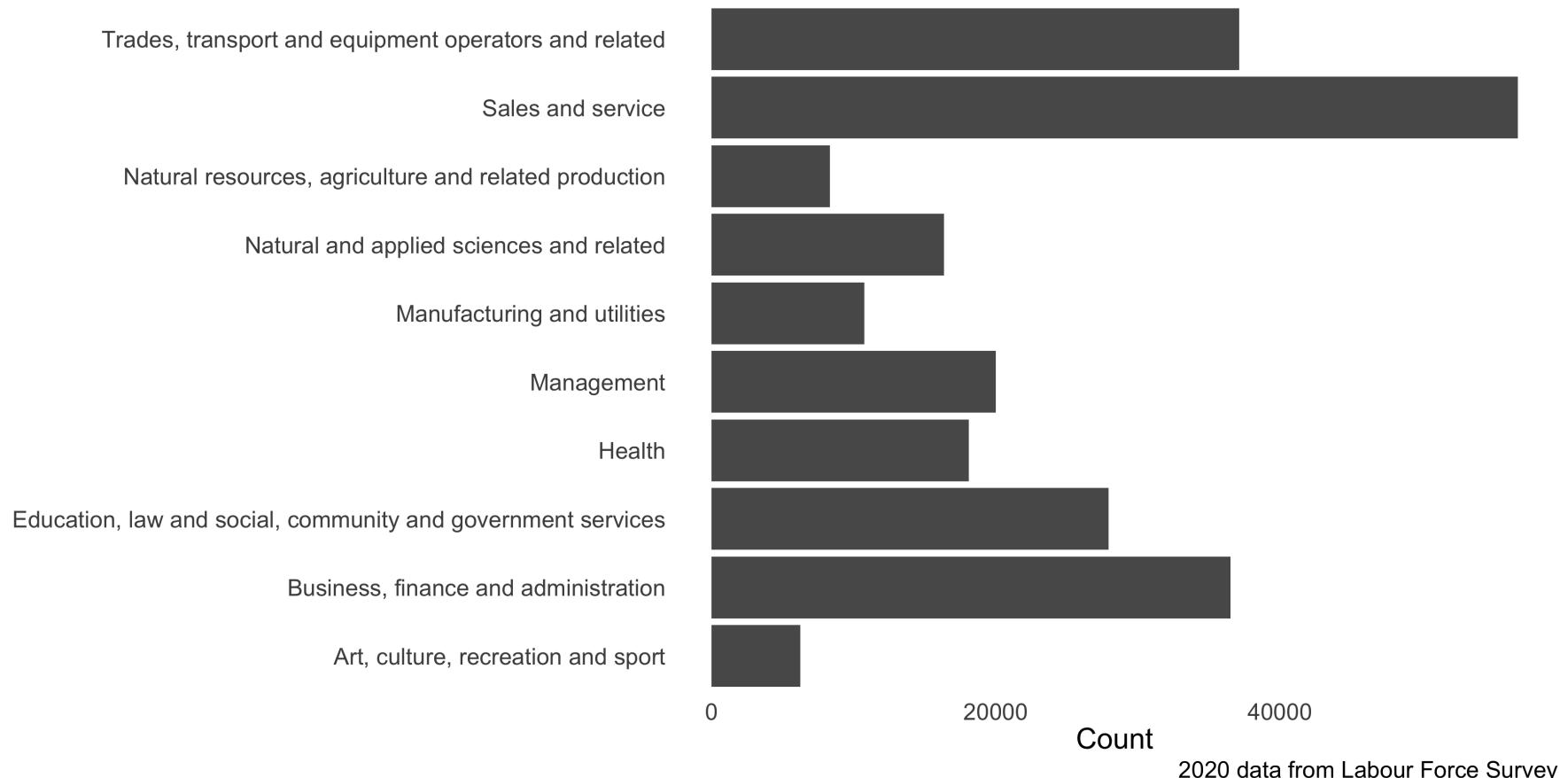
Describing the data & Telling a story

Counts, Distributions & Comparing Variables

Abstracting Further from the Data

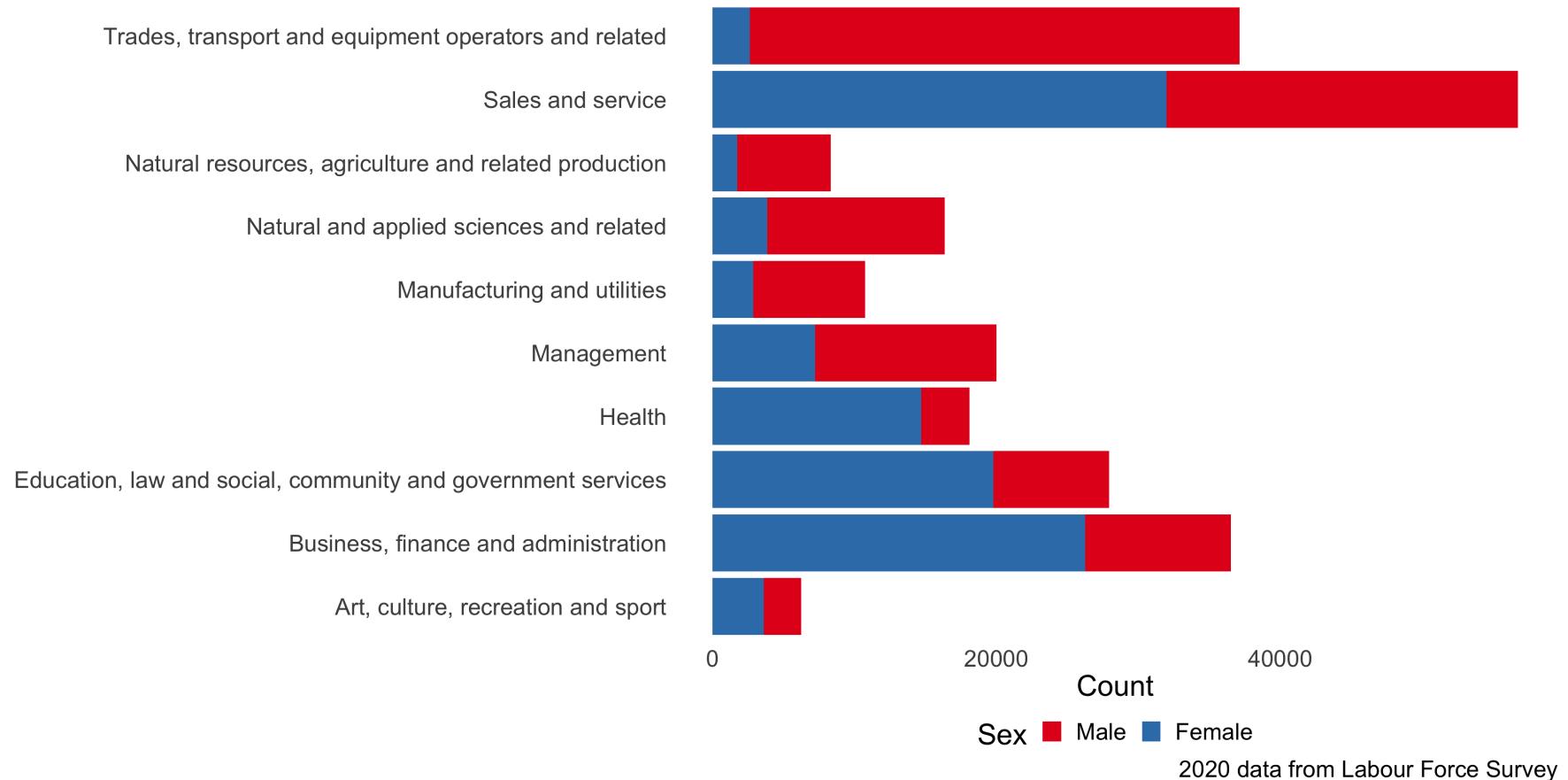
Counts

Employment by Occupation



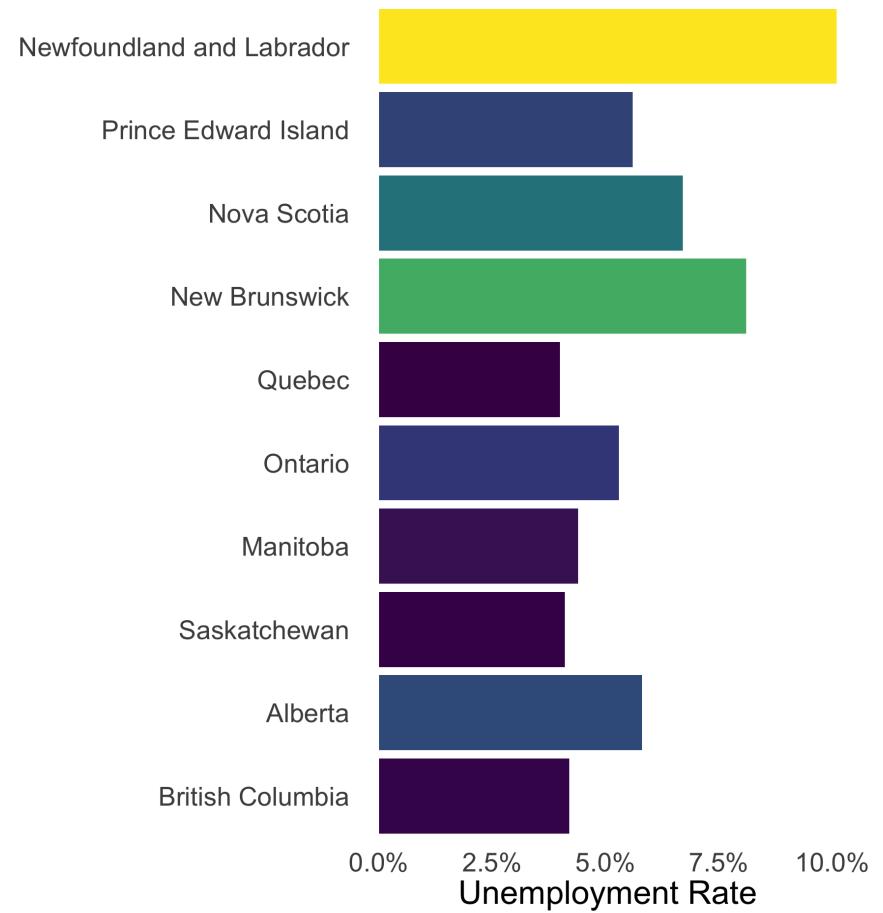
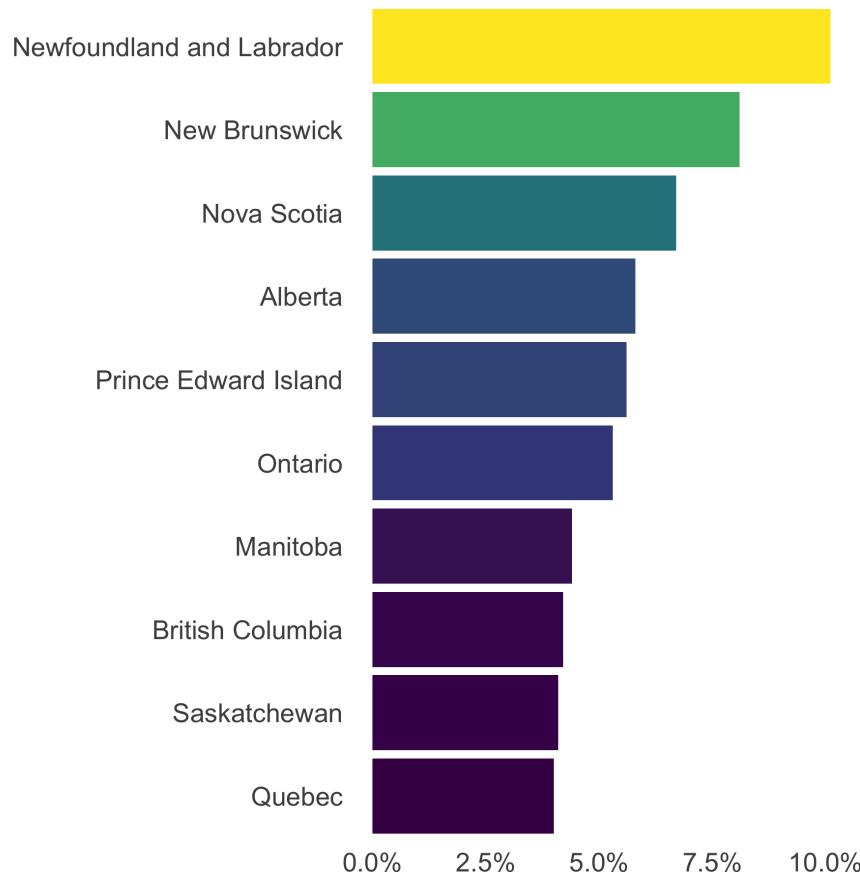
2020 data from Labour Force Survey

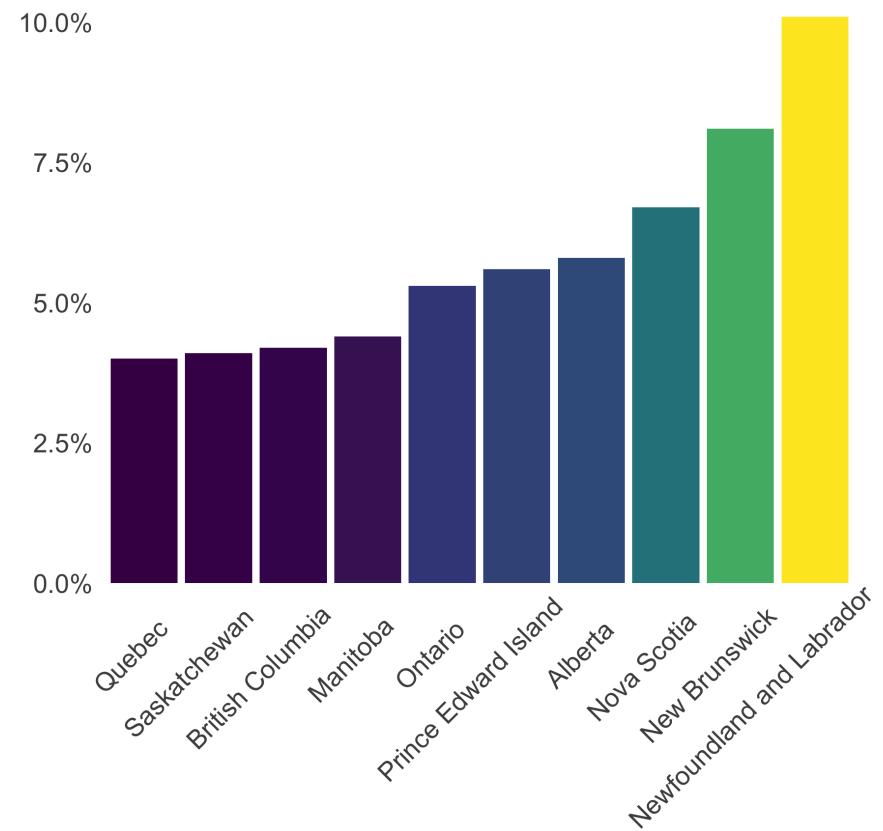
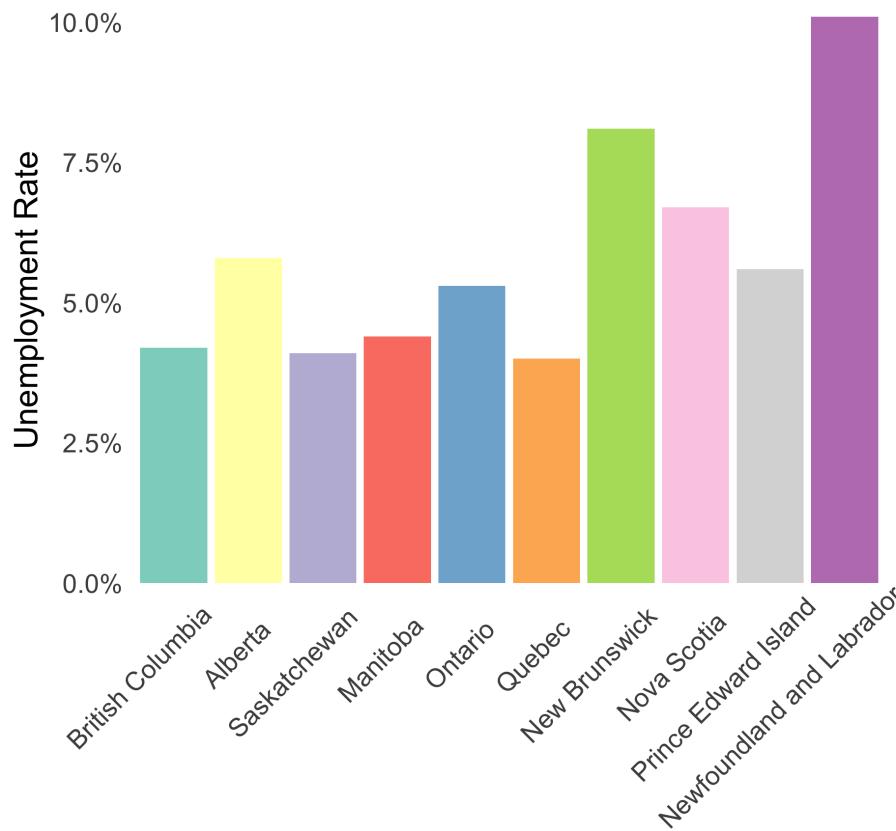
Employment by Occupation



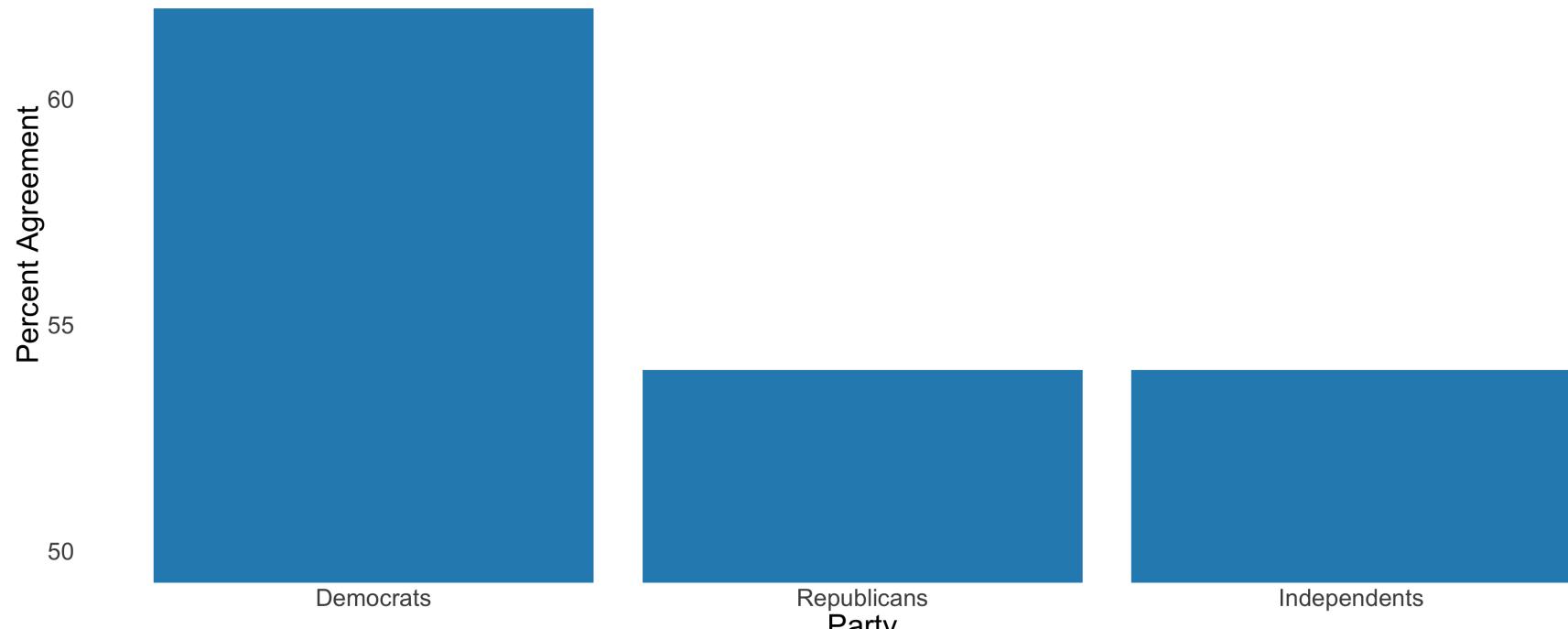
Sex ■ Male ■ Female

2020 data from Labour Force Survey



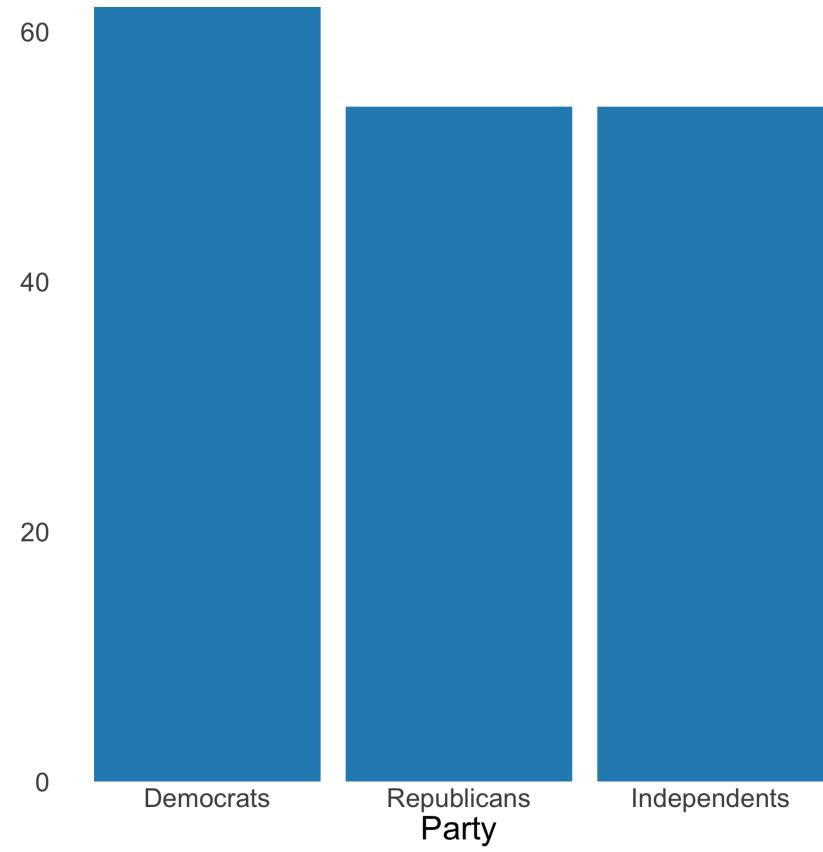
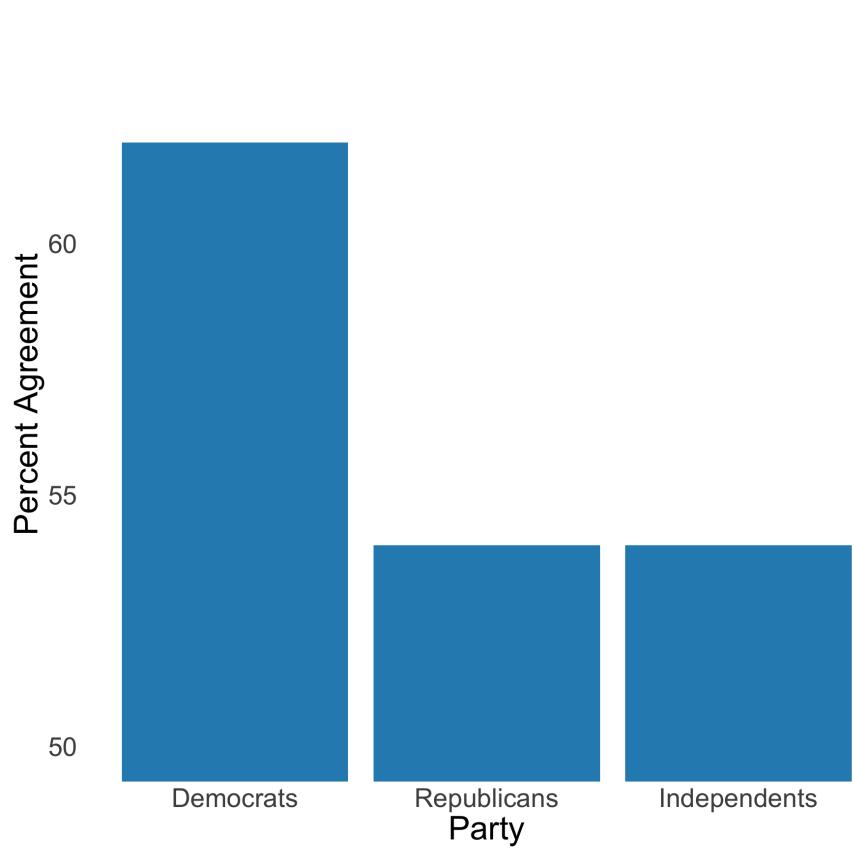


Terri Schiavo case



Originally published by CNN

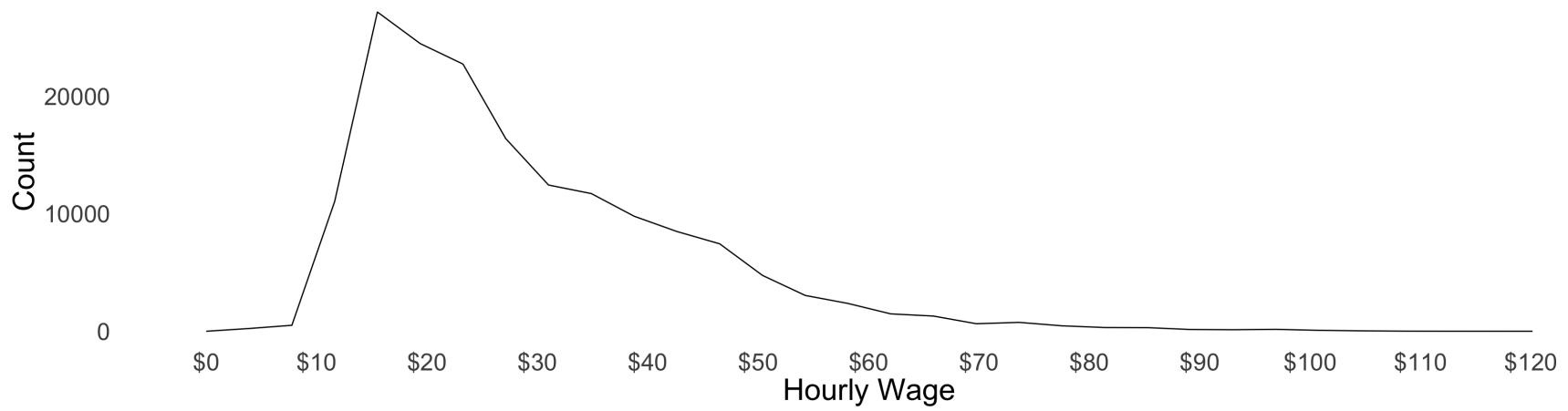
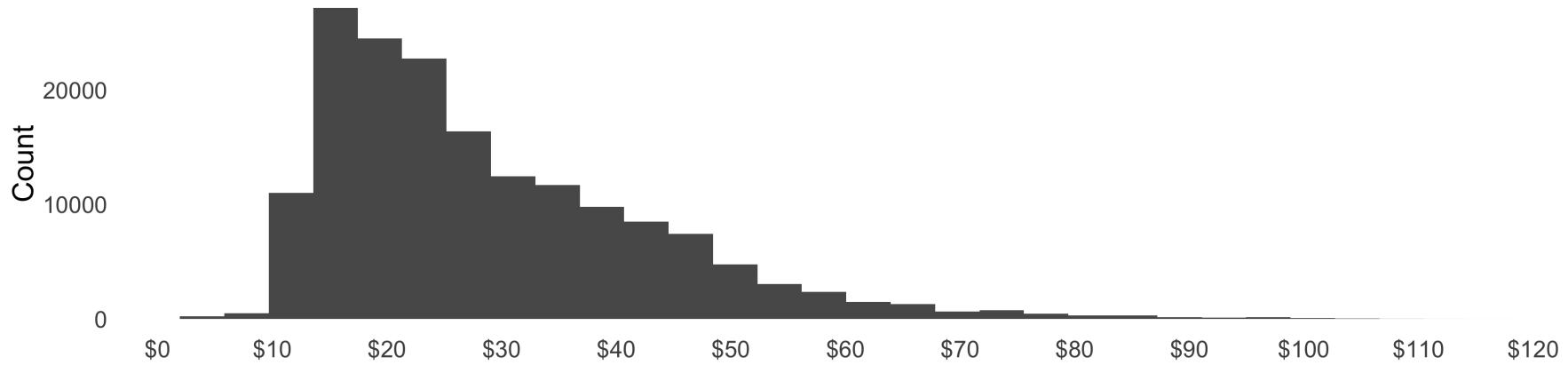
Source: <https://www.statisticshowto.com/probability-and-statistics/descriptive-statistics/misleading-graphs/>

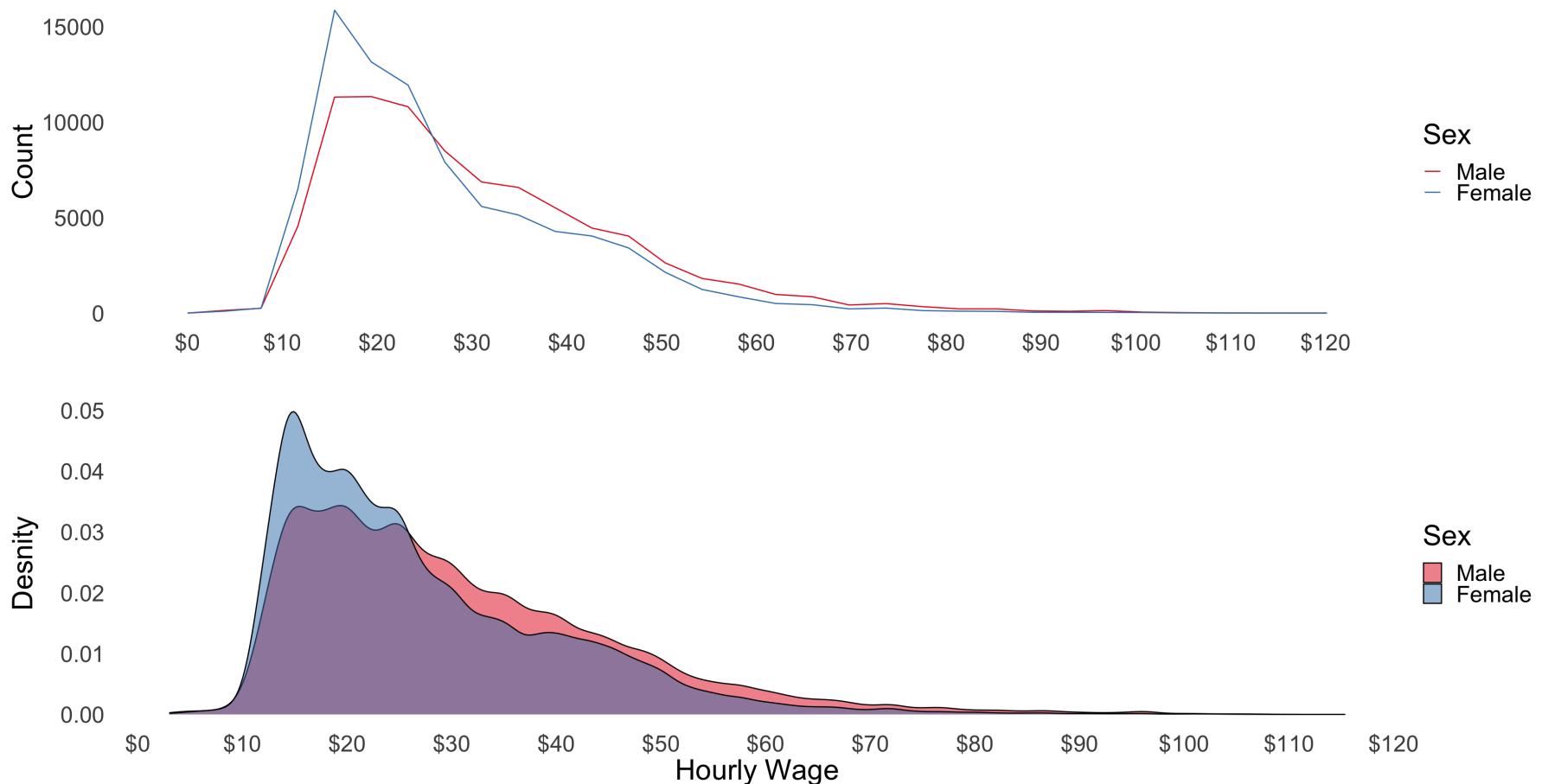


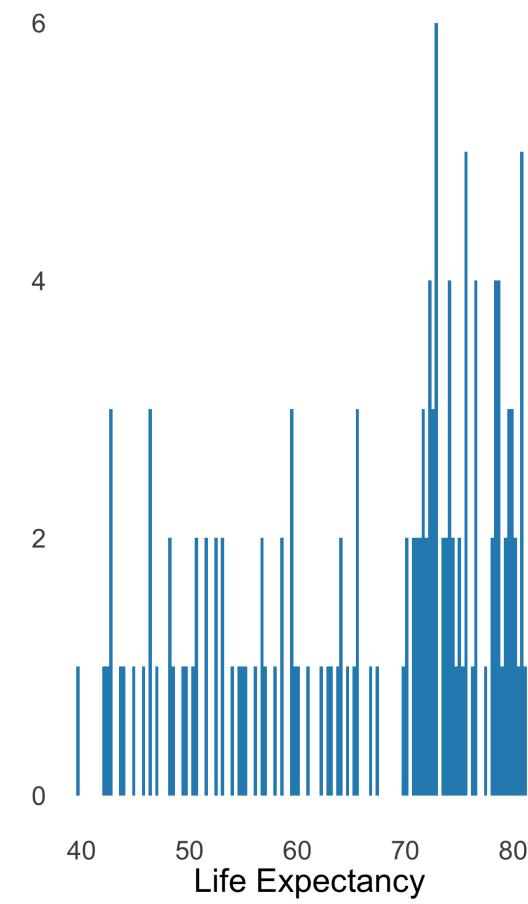
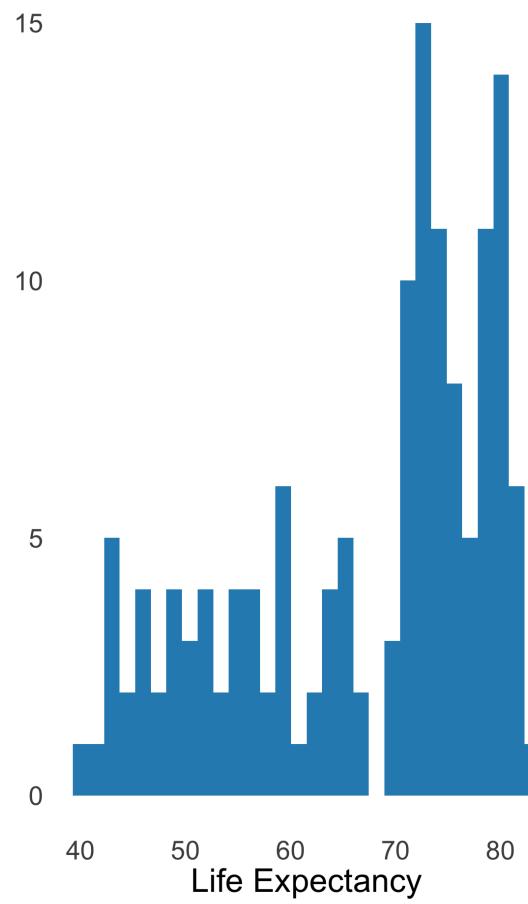
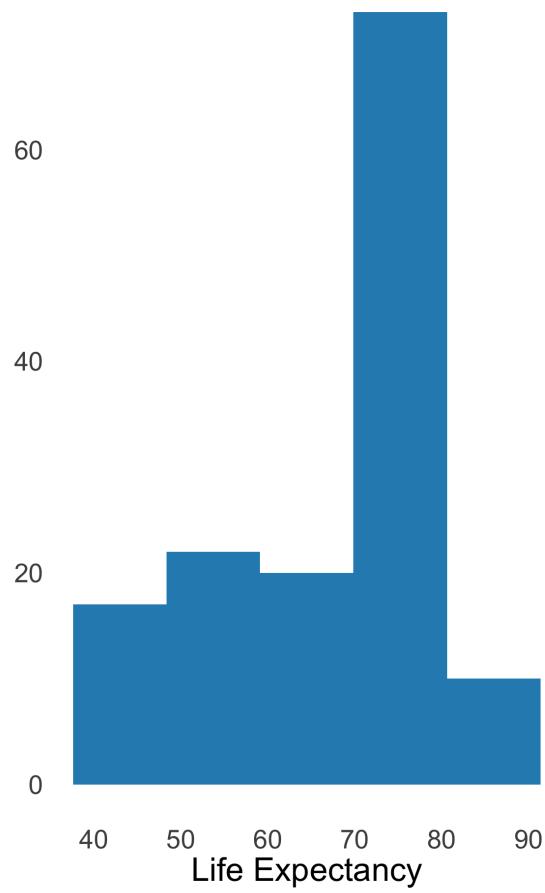


Source: https://junkcharts.typepad.com/junk_charts/2013/05/more-power-brings-more-responsibility.html

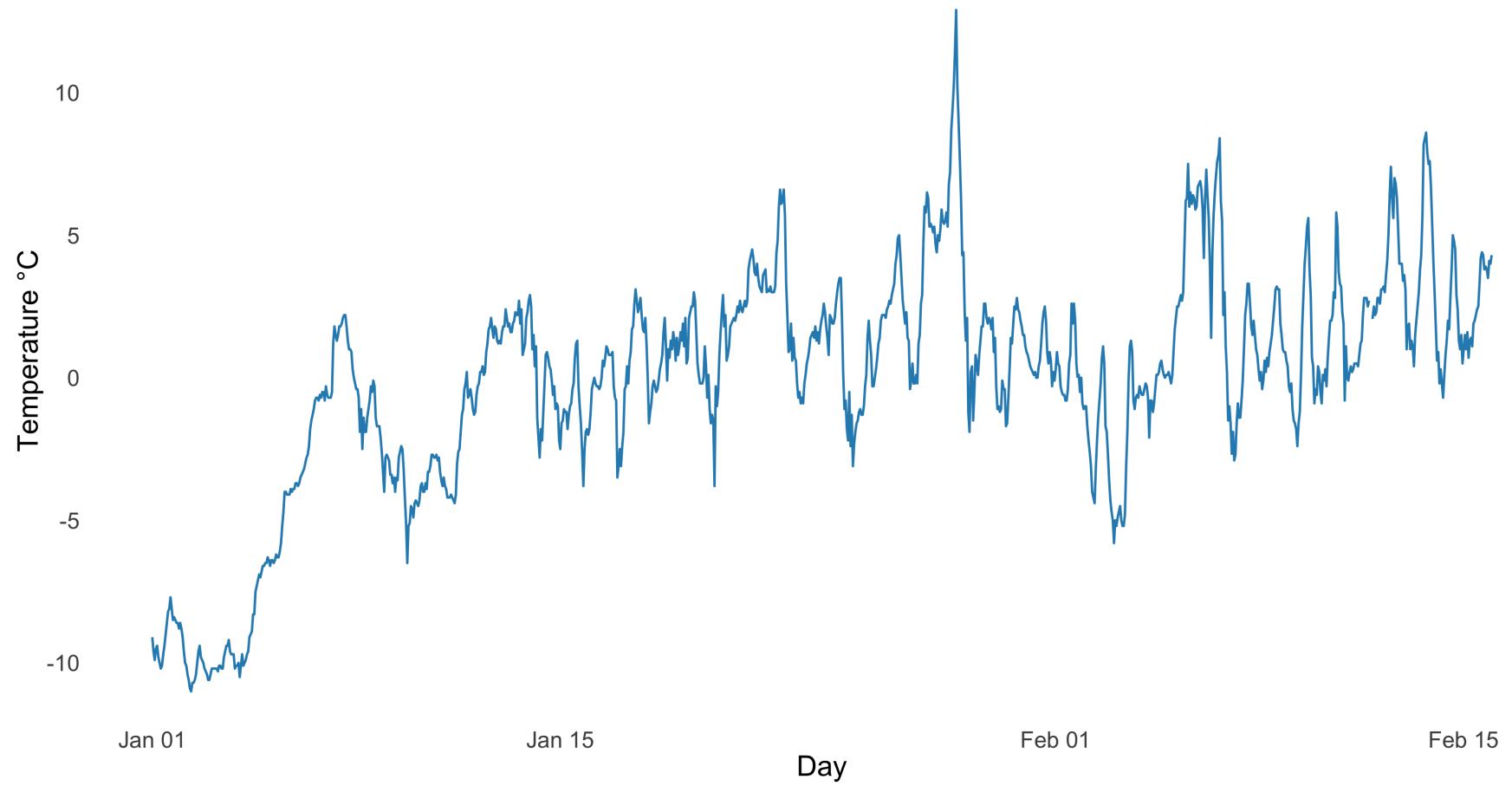
Distributions

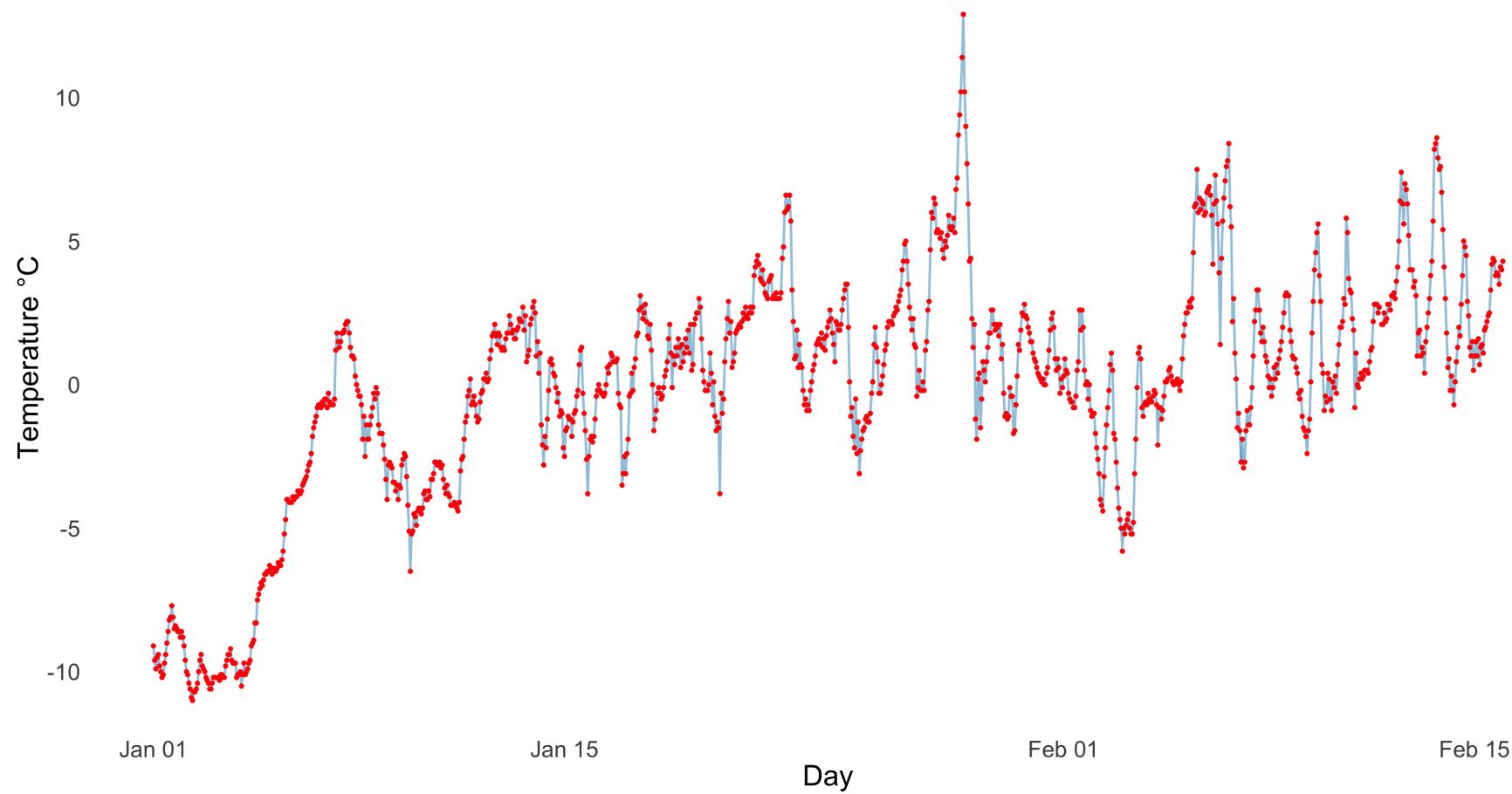


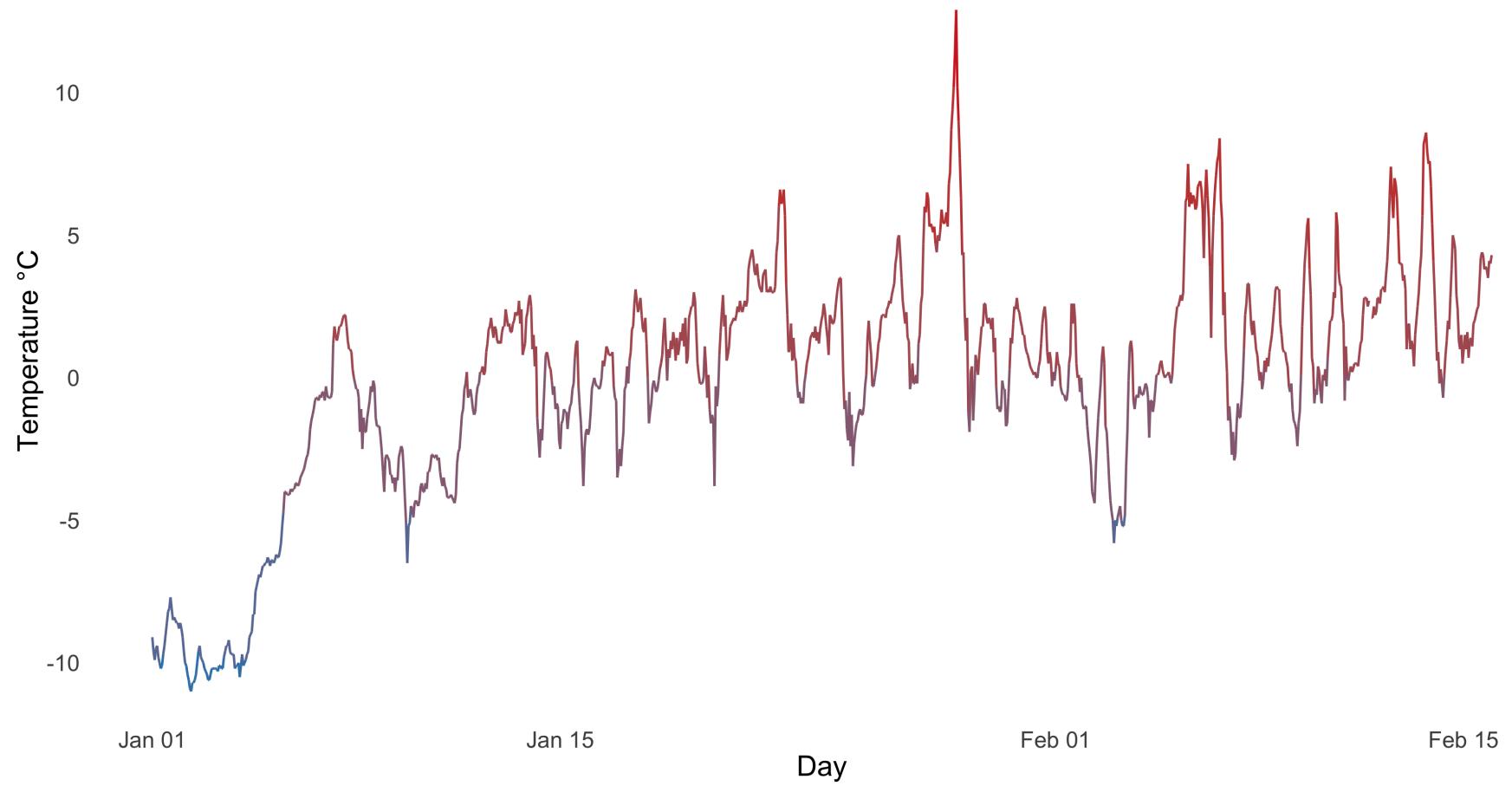


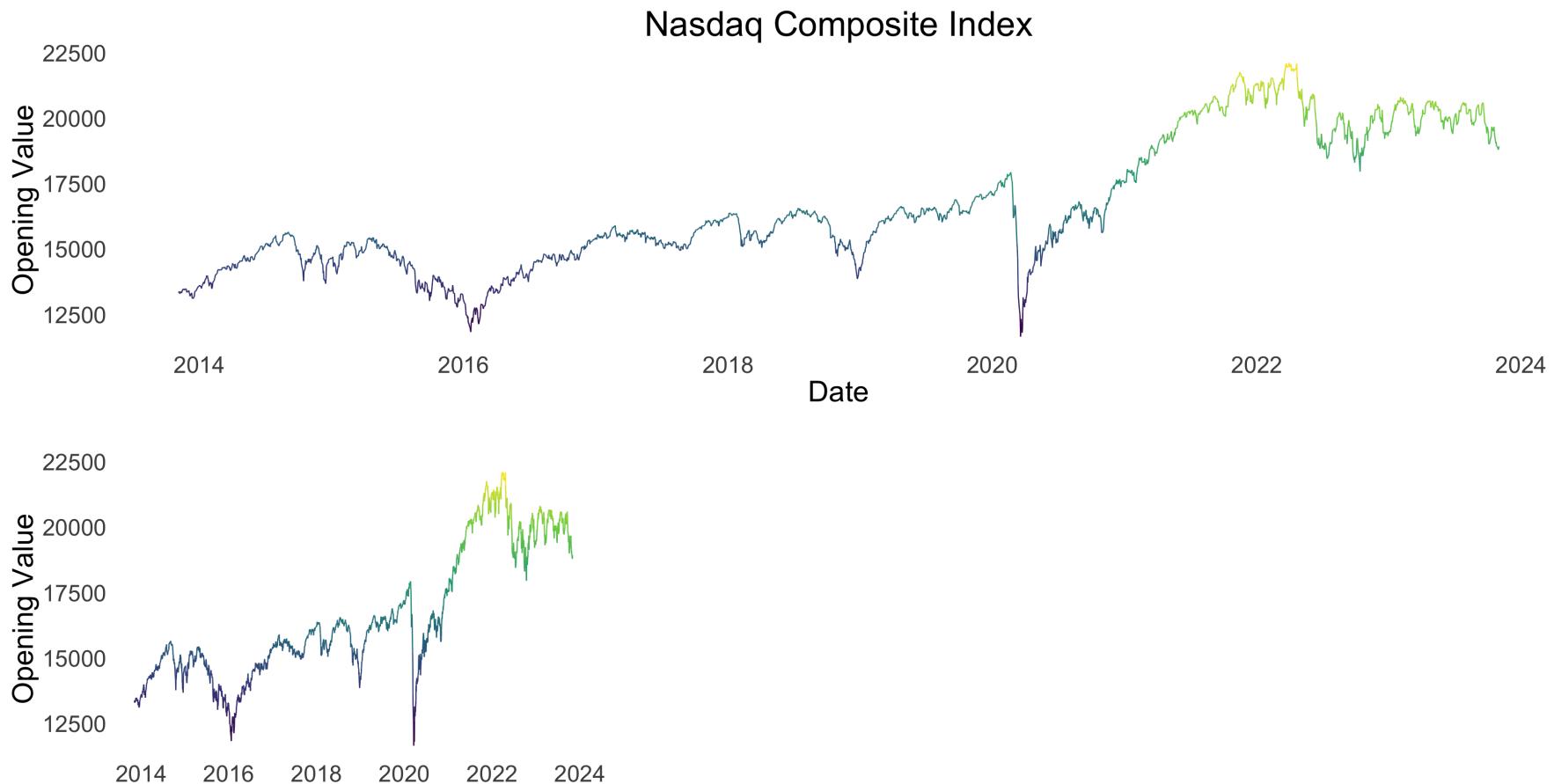


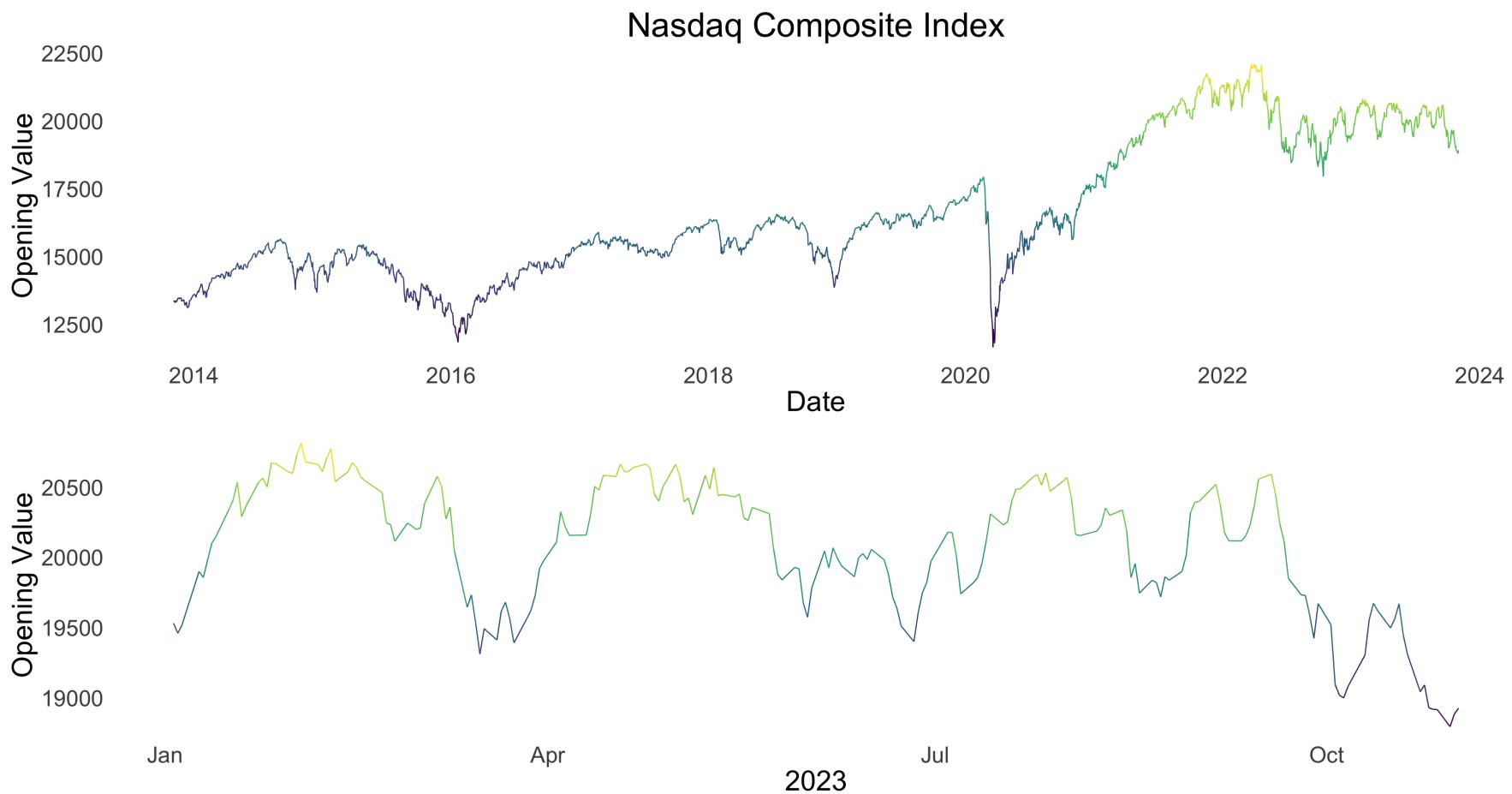
Time Series



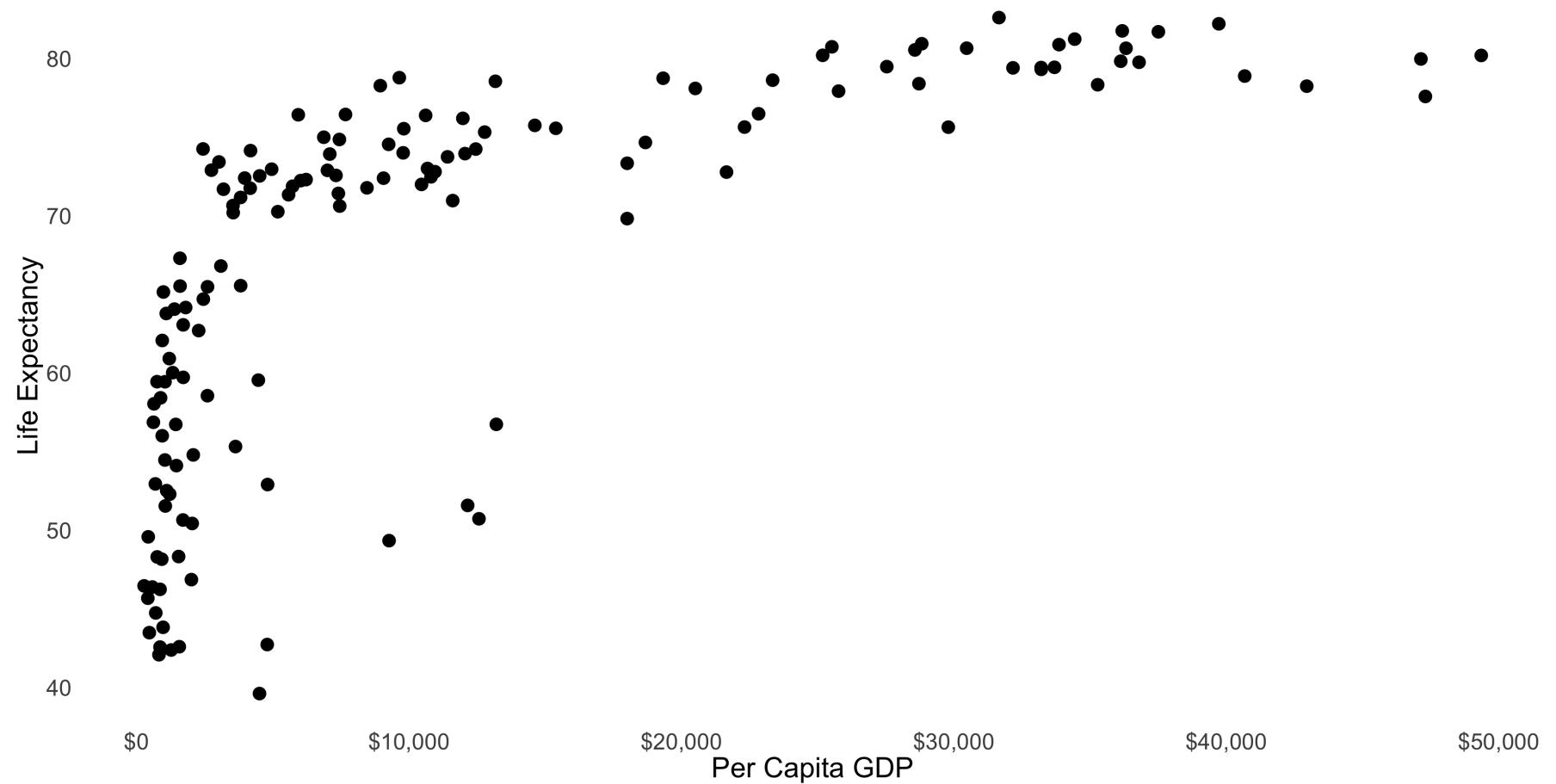


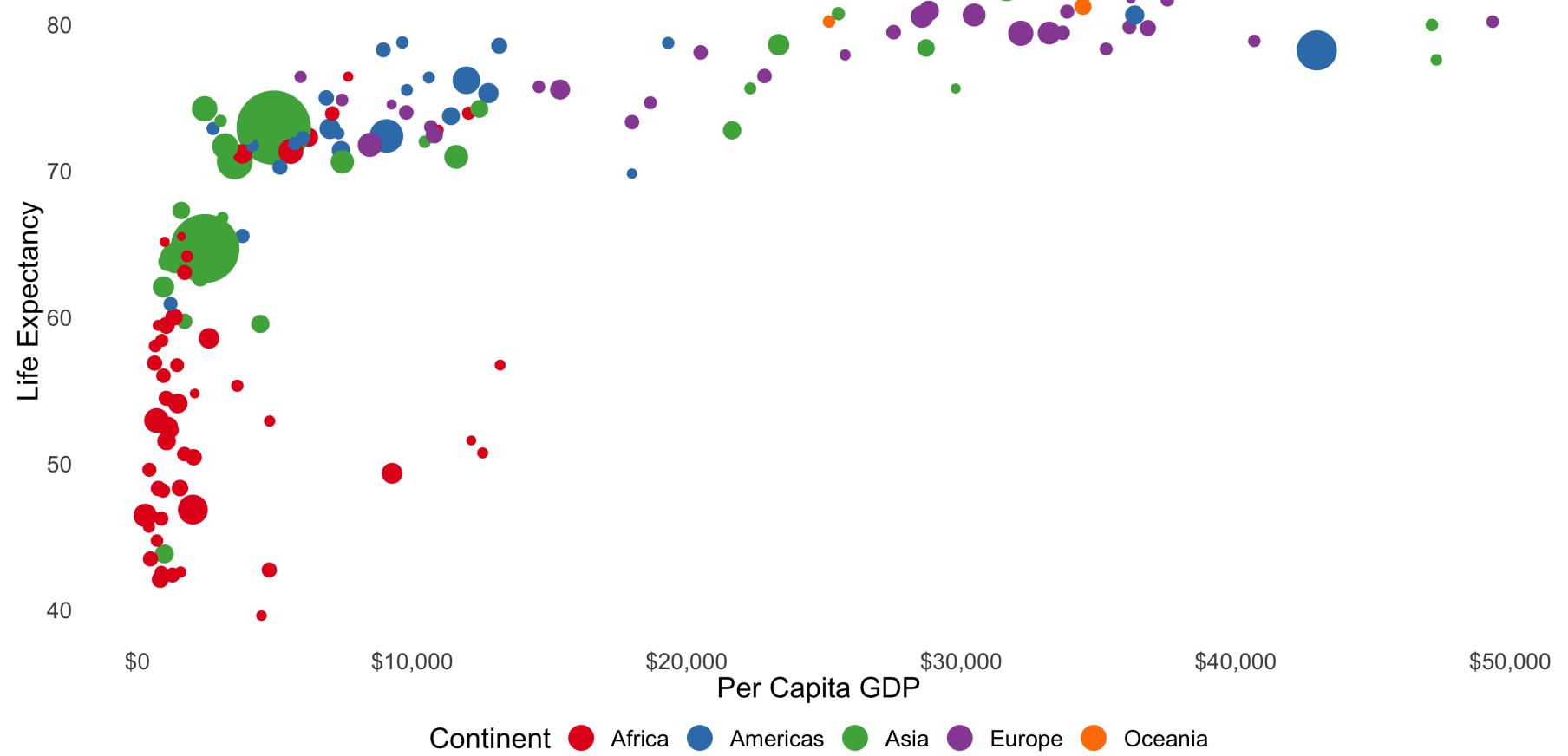


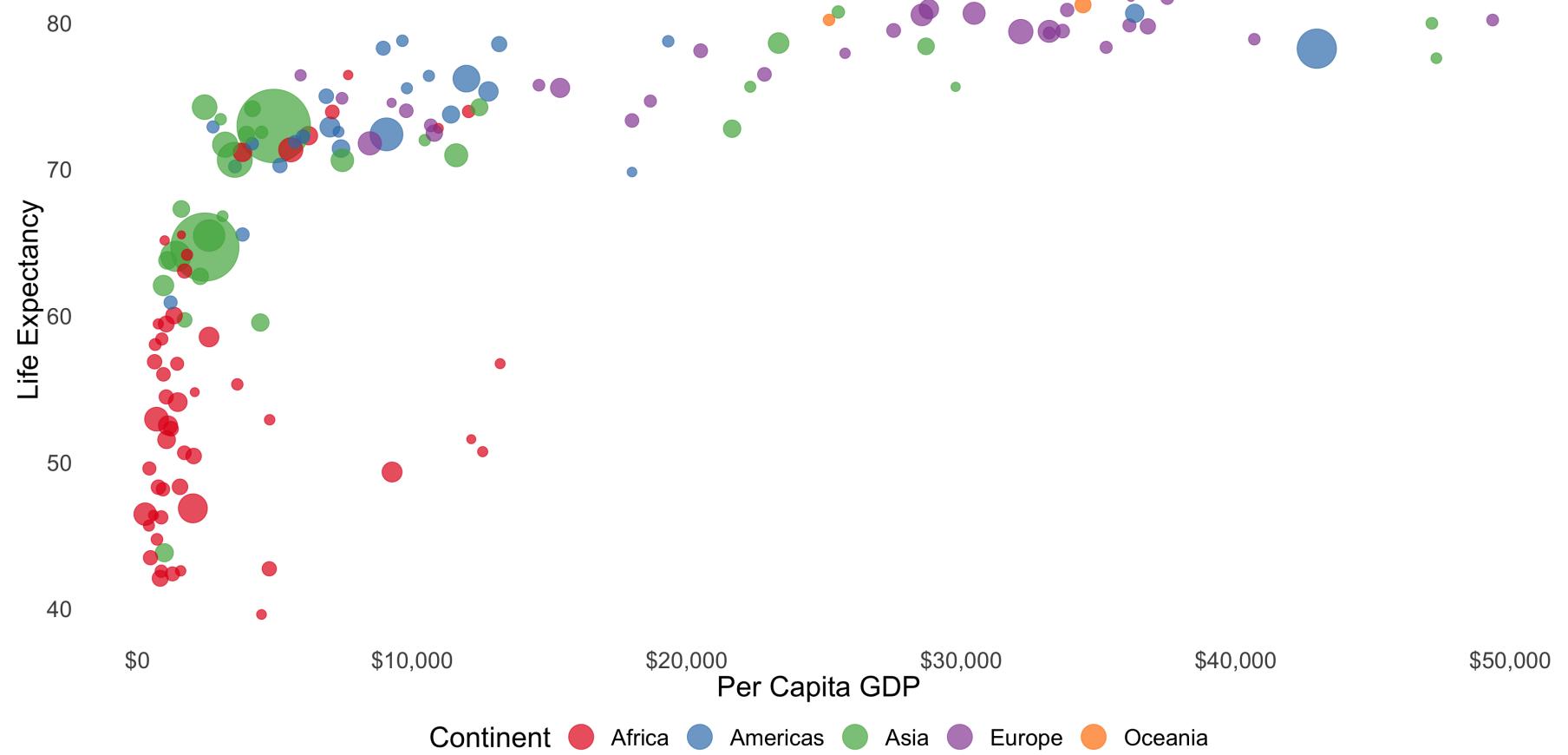


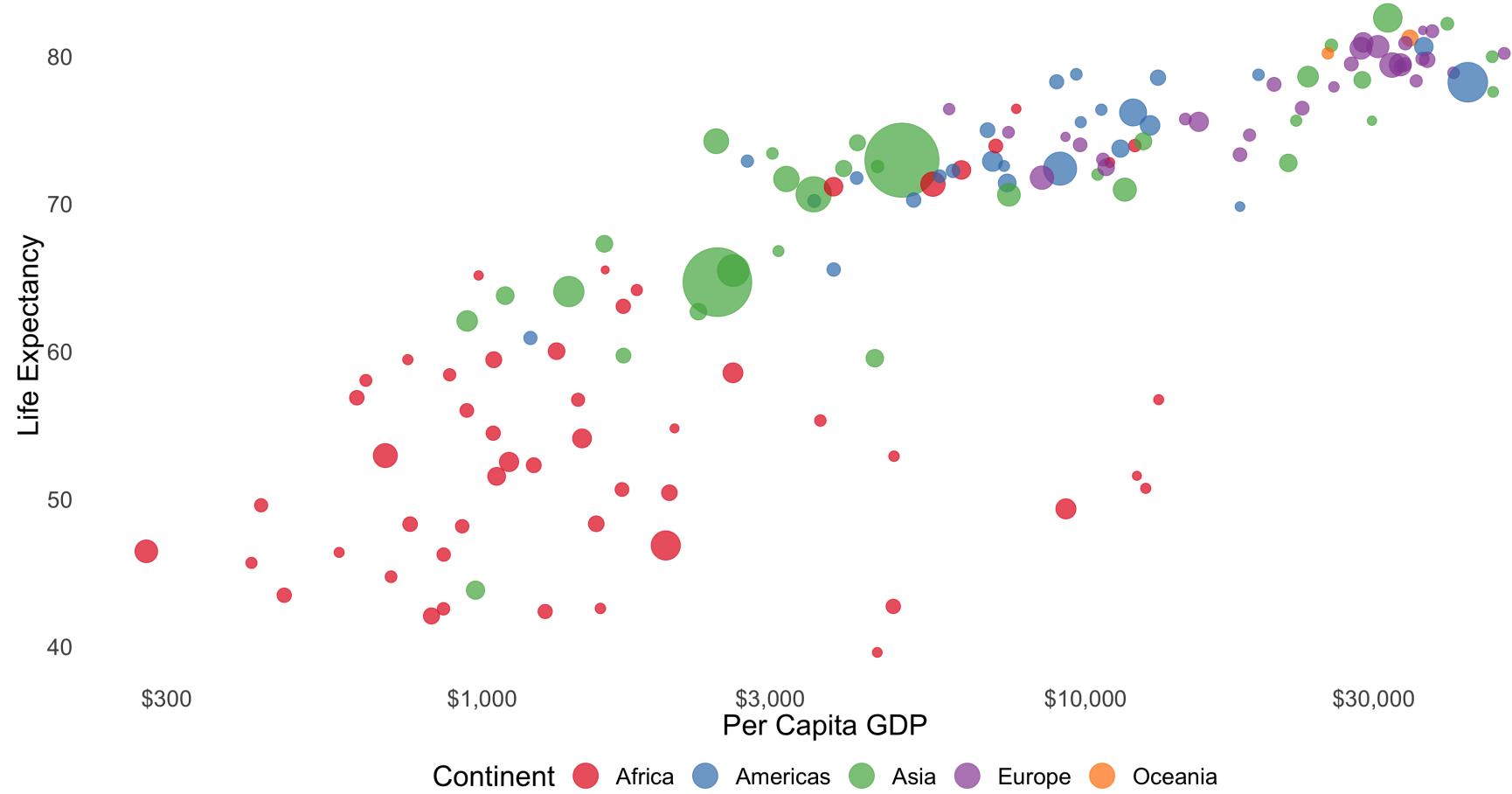


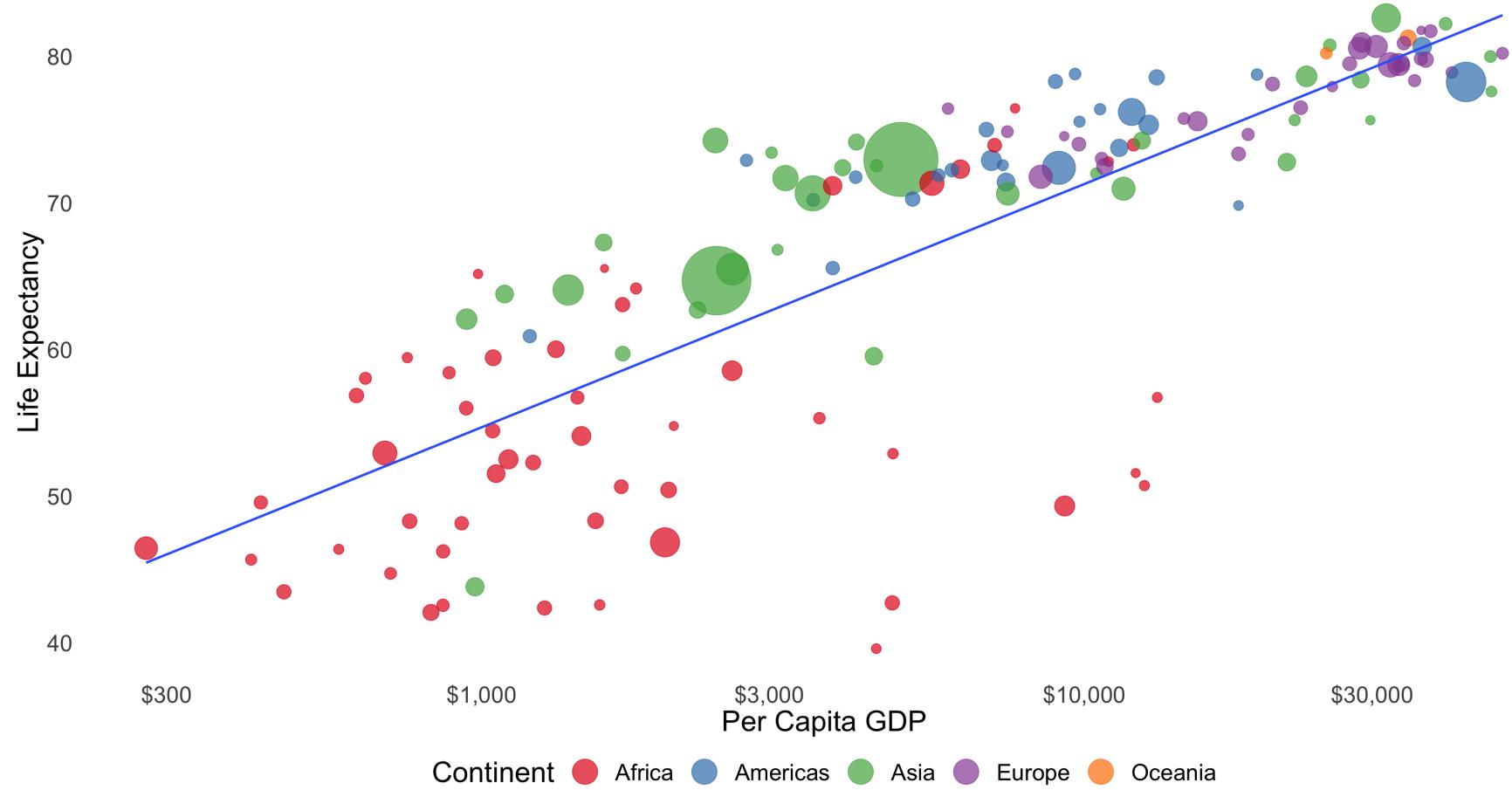
Comparing Variables

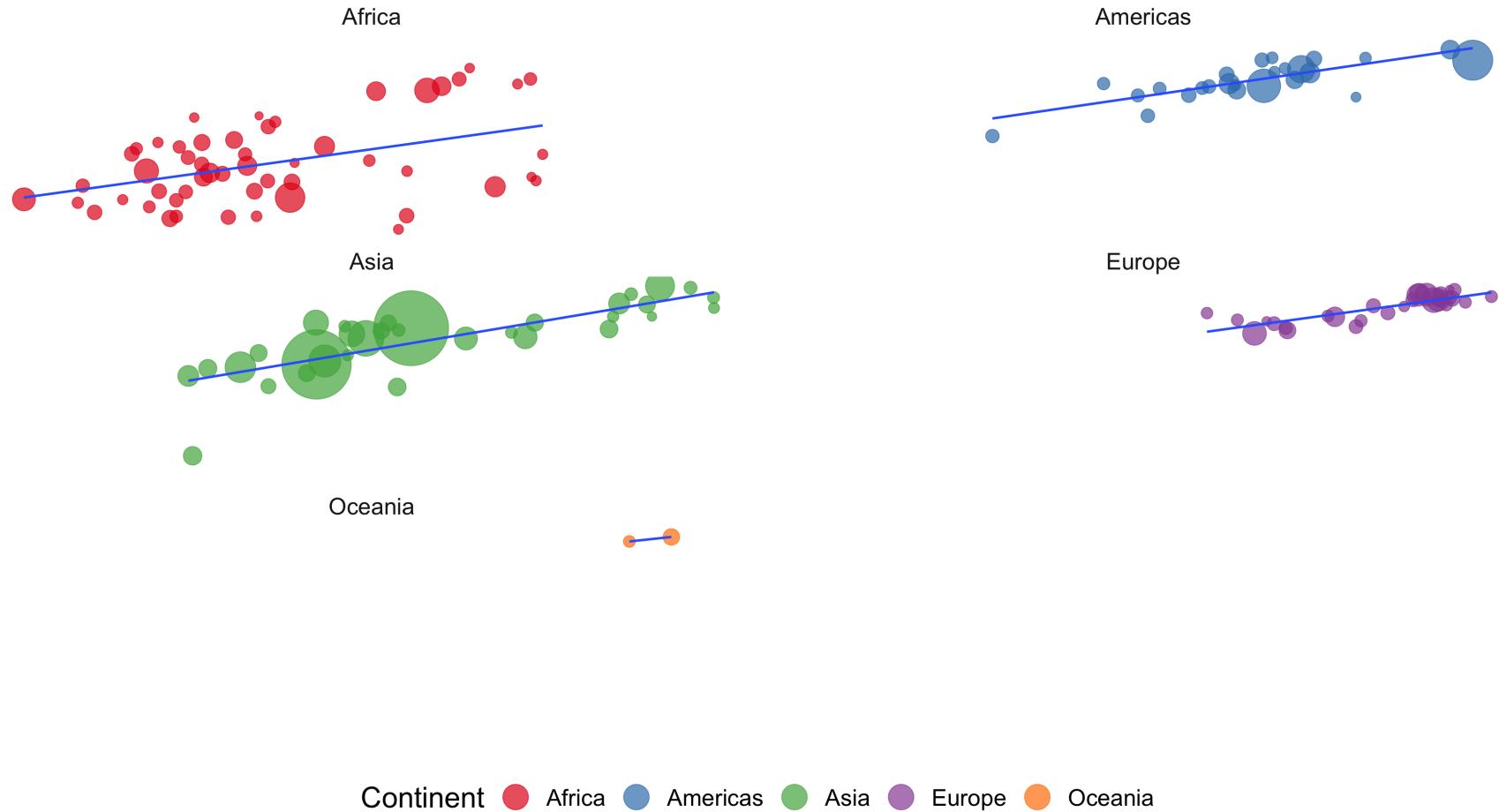












Know Your Audience

A Data Table May be Enough

Show 10 entries Search:

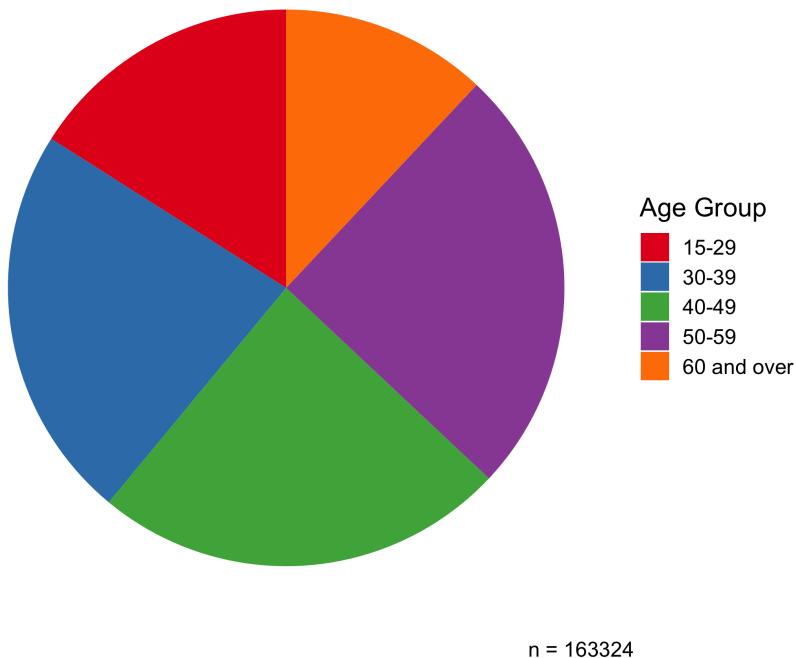
Average Life Expectancy	Number of Countries
40	1
42	2
43	4
44	1
45	1
46	4
47	1
48	3
49	1
50	2

Showing 1 to 10 of 40 entries

Previous 2 3 4 Next

Or a Pie Chart

Full-time employment by age group

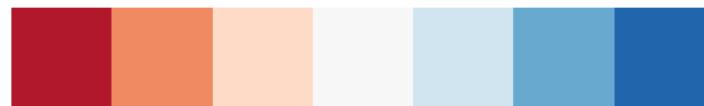


Colour Should be Meaningful

Sequential



Diverging



Qualitative



Colour Blindness

