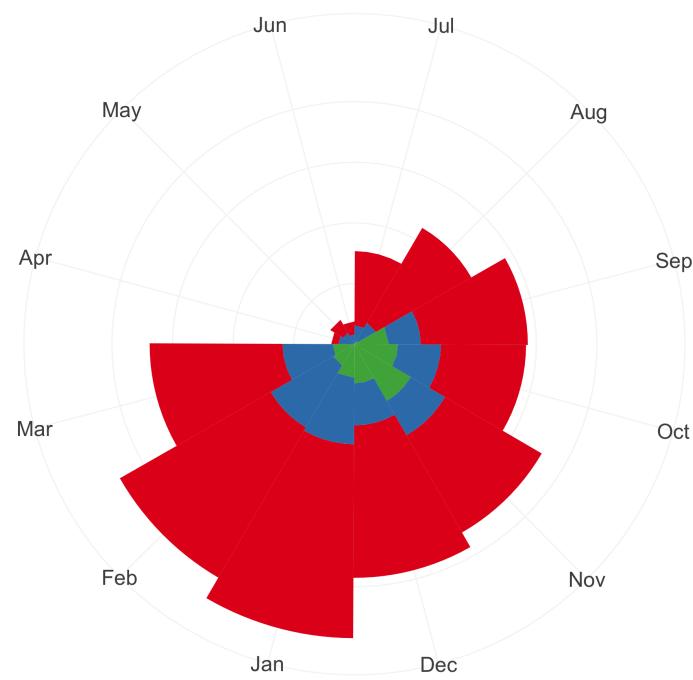


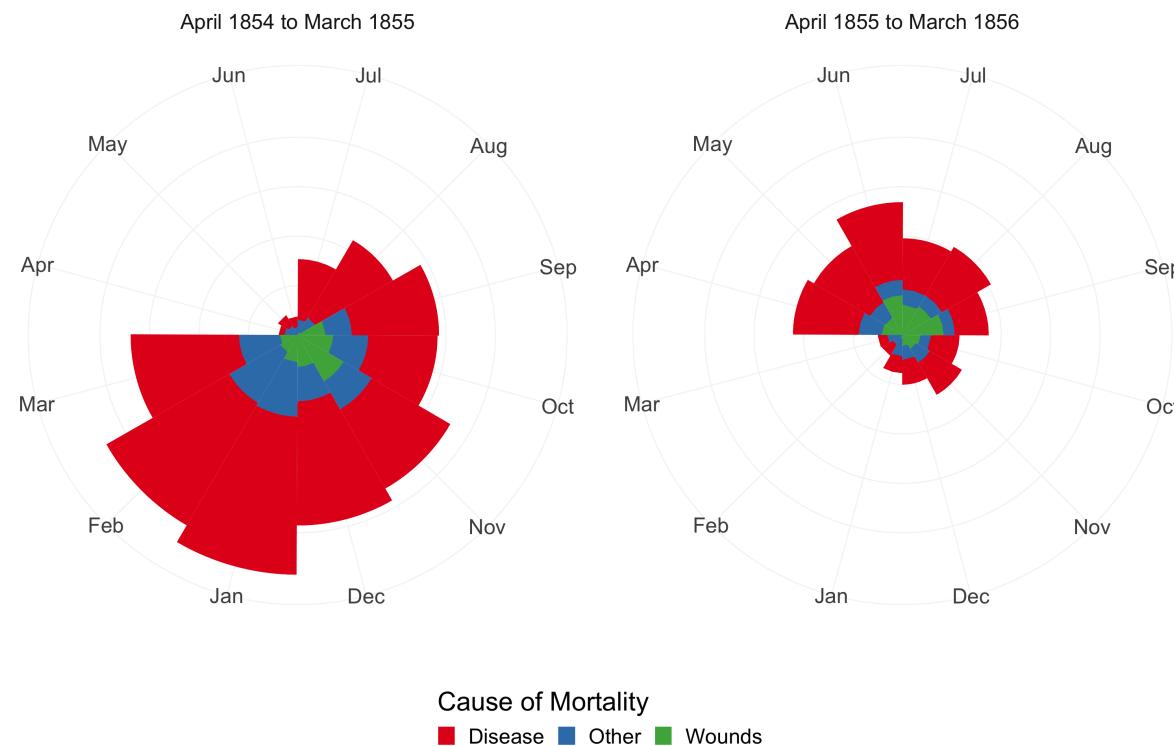
Data Visualization

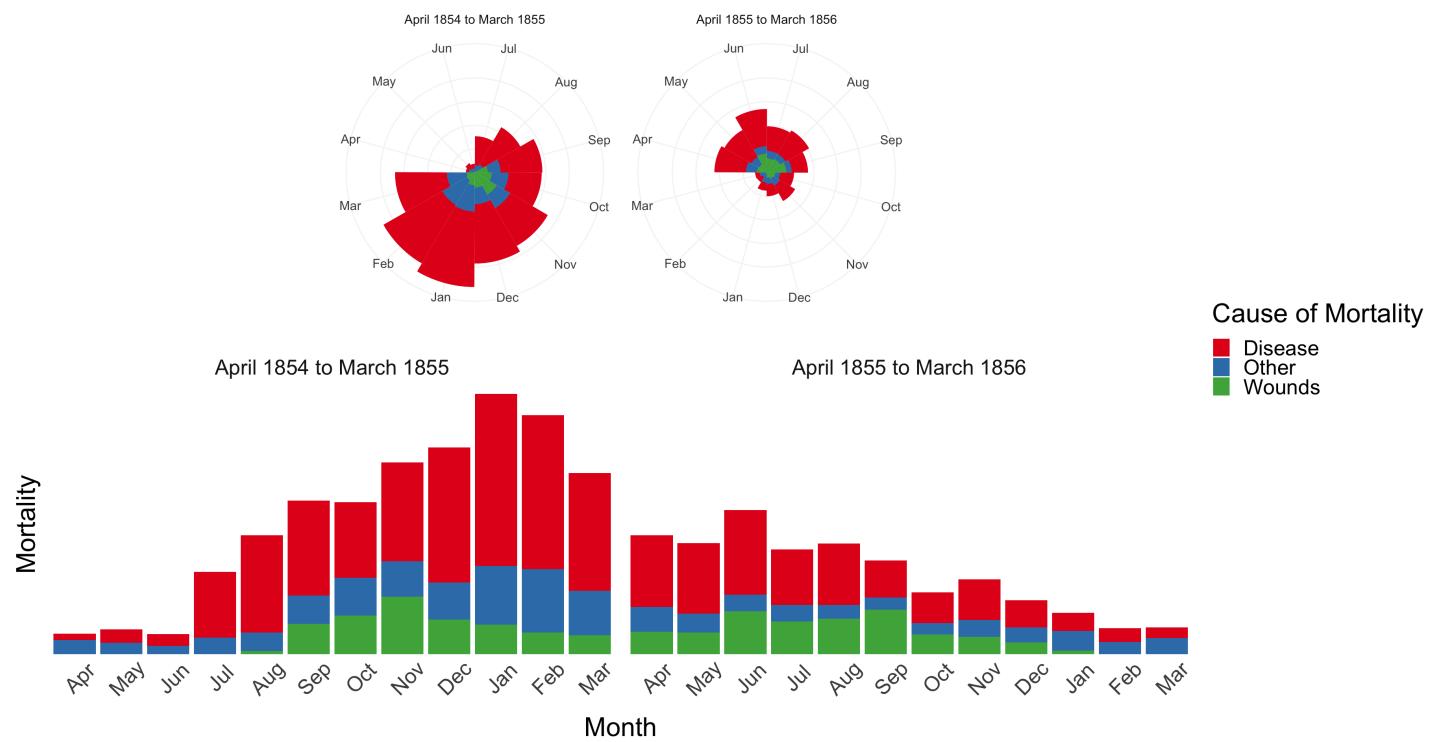
CORH 205

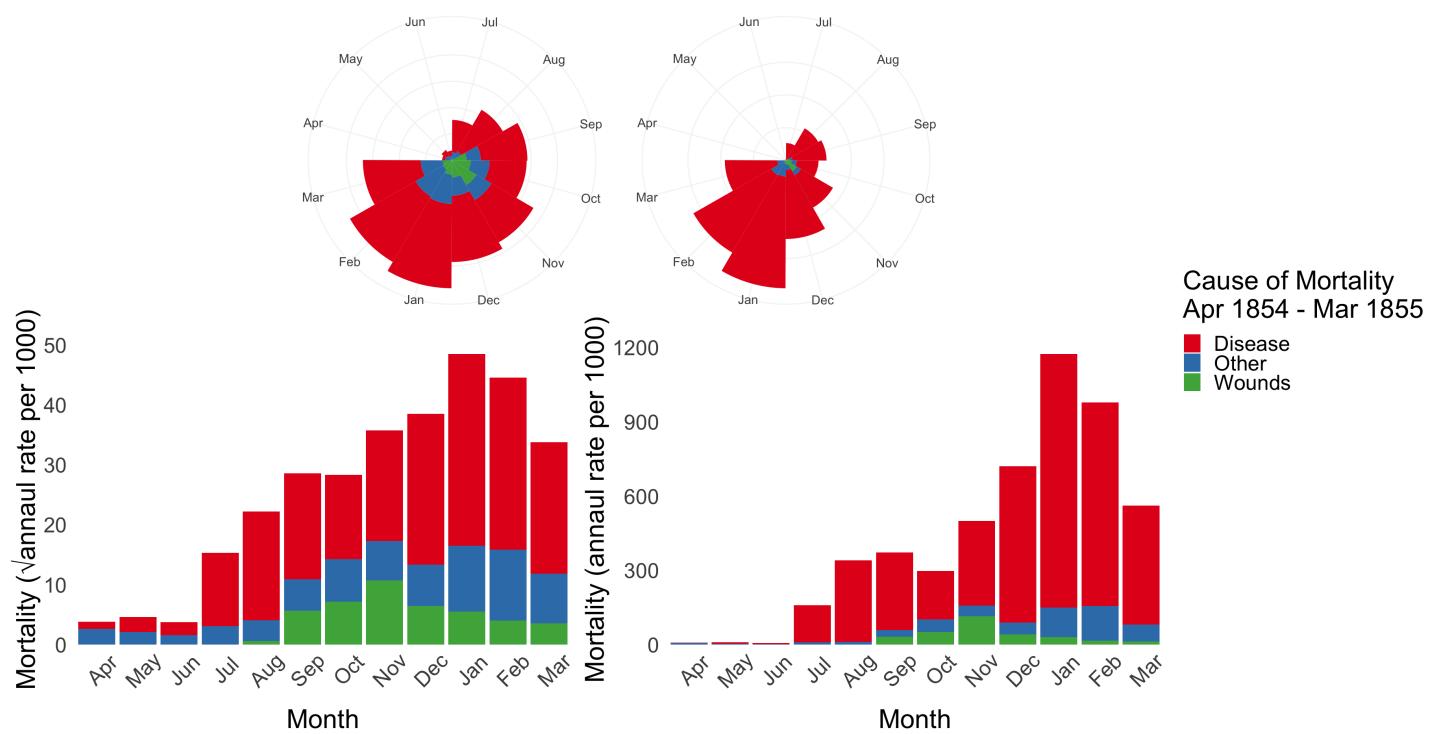
Mathew Vis-Dunbar

March 2023









Overview

- Data visualization core concepts
 - What and why
- Data types and common visualizations
 - Categorical and numeric data
 - Common issues / considerations
- Layering variables
- Statistical computations on data
- Data quality and bias

What is Data Visualization?

What is Data Visualization?

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

What is Data Visualization?

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

This involves abstraction as shapes, colours etc are used represent the data.

What is Data Visualization?

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

This involves abstraction as shapes, colours etc are used represent the data.

Visual and data literacies are needed to interpret both the data and the abstraction.

Why Do We Visualize Data?

A Definition

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](#).

Attributes & Perception

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](#).

7	5	9	1	5	3	3	9	7	2	3	3
4	4	8	8	7	6	9	5	3	5	8	9
8	3	9	9	1	9	3	3	4	5	5	7
4	3	9	5	6	7	6	7	4	2	3	8
4	8	1	3	6	9	5	1	6	2	2	5
6	5	4	4	9	9	9	2	8	2	4	7
8	1	7	8	8	9	9	5	6	9	1	7
5	9	3	9	1	4	2	5	9	2	5	9
1	6	8	1								

7	5	9	1	5	3	3	9	7	2	3	3
4	4	8	8	7	6	9	5	3	5	8	9
8	3	9	9	1	9	3	3	4	5	5	7
4	3	9	5	6	7	6	7	4	2	3	8
4	8	1	3	6	9	5	1	6	2	2	5
6	5	4	4	9	9	9	2	8	2	4	7
8	1	7	8	8	9	9	5	6	9	1	7
5	9	3	9	1	4	2	5	9	2	5	9
1	6	8	1								

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								

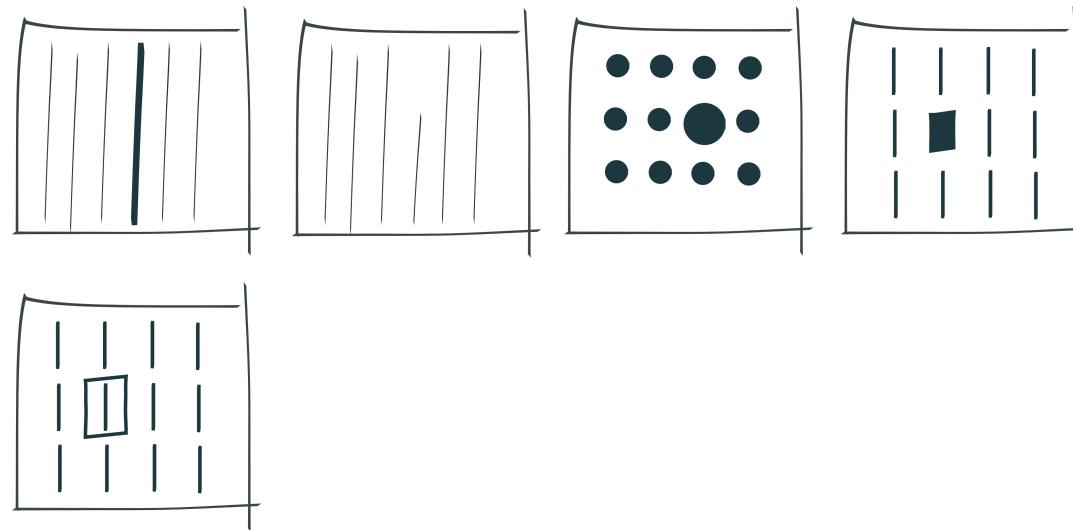
E	5	9	A	5	C	C	9	E	2	C	C
4	4	8	8	E	6	9	5	C	5	8	9
8	C	9	9	A	9	C	C	4	5	5	E
4	C	9	5	6	E	6	E	4	2	C	8
4	8	A	C	6	9	5	A	6	2	2	5
6	5	4	4	9	9	9	2	8	2	4	E
8	A	E	8	8	9	9	5	6	9	A	E
5	9	C	9	A	4	2	5	9	2	5	9
A	6	8	A								

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								

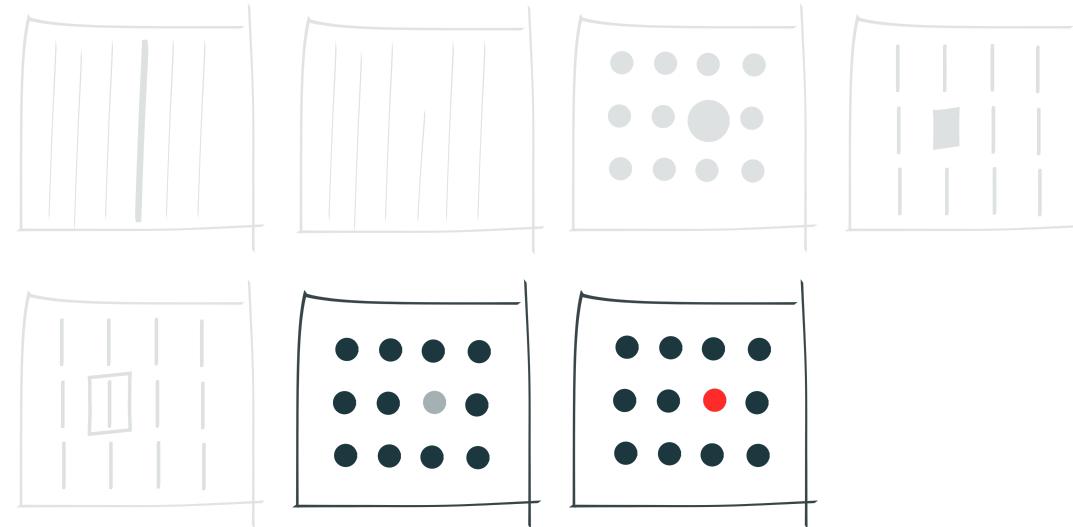
Preattentive Attributes

- Form
- Colour
- Position

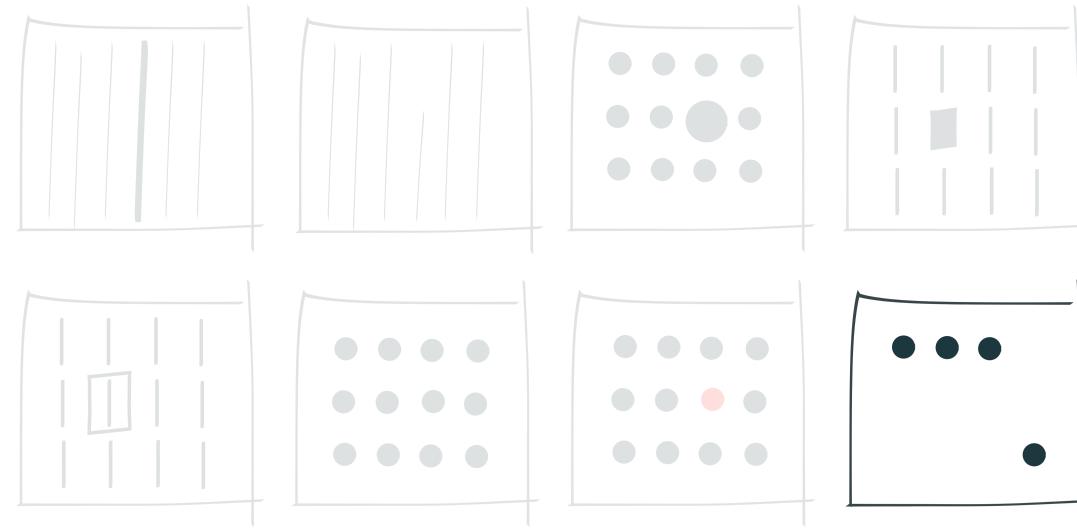
Form



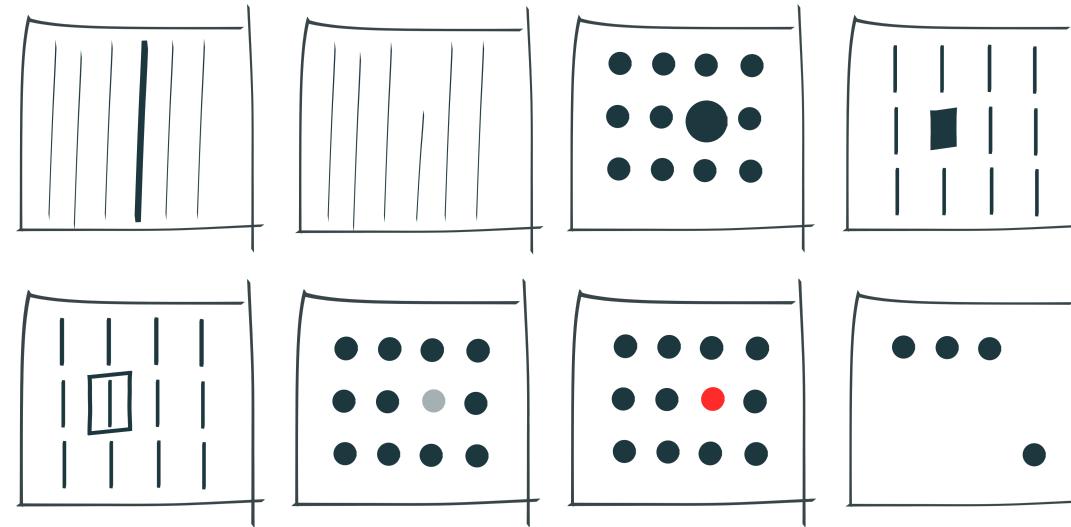
Colour



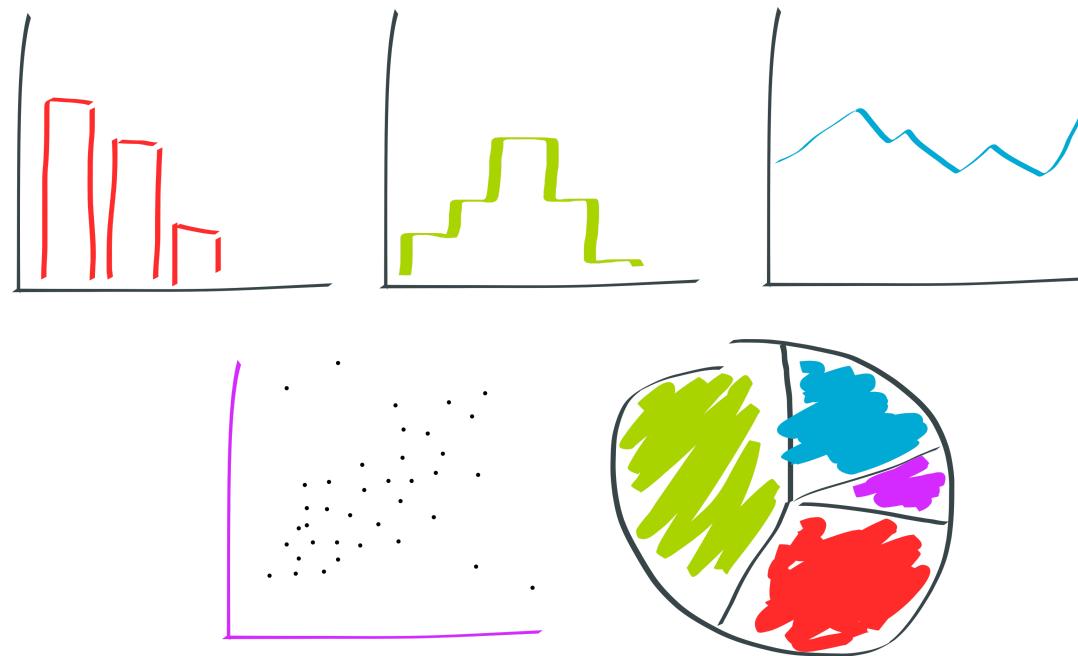
Position

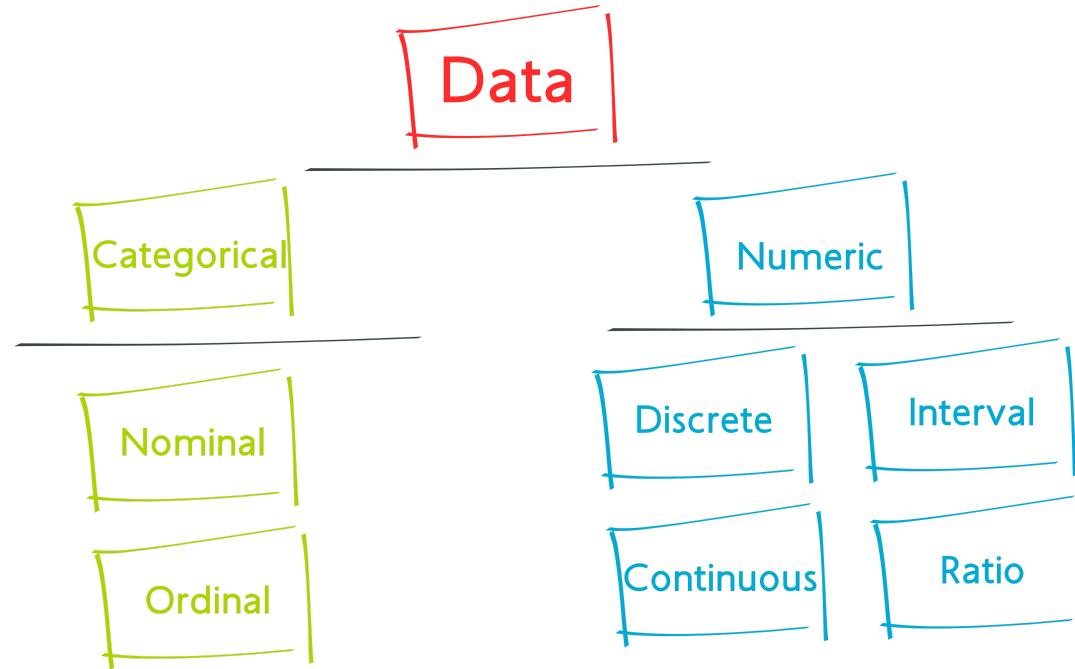


Preattentive Attributes



Types of Visualisations





Categorical Data

Nominal data

No order



Ordinal data

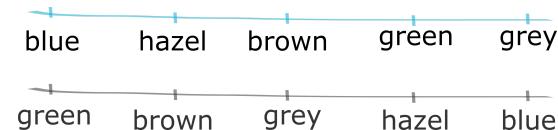
Intrinsic order



Categorical Data

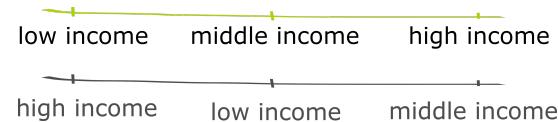
Nominal data

No order



Ordinal data

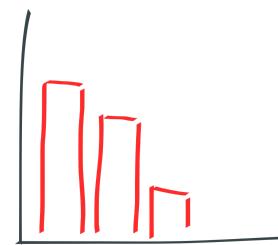
Intrinsic order



Bar Plots

Visualization tool

Frequency plots with bar charts

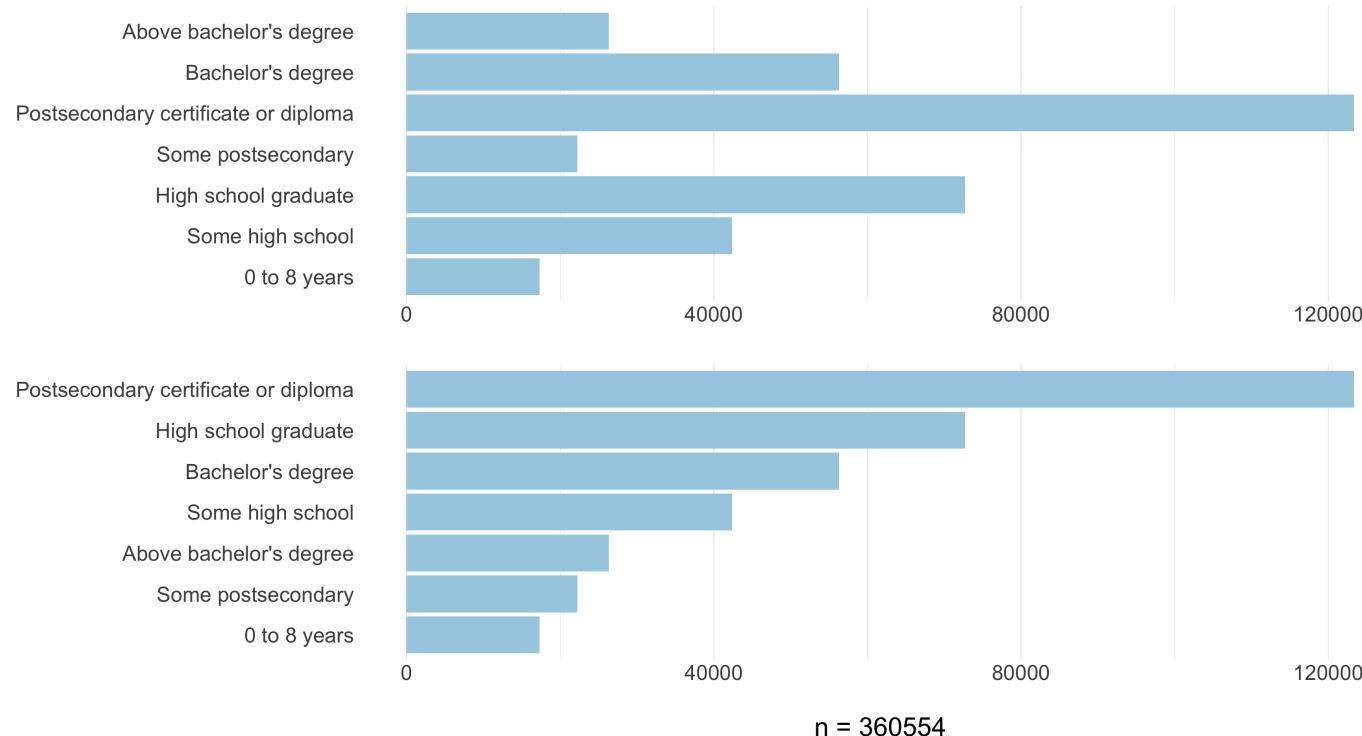


variable	class	first_values
Record.number	integer	1, 2, 3, 4, 5, 6
Survey.year	integer	2020, 2020, 2020, 2020, 2020, 2020
Survey.month	character	January, January, January, January, January, January
Labour.force.status	character	Not in labour force, Employed, at work, Not in labour force, Employed, at work, Employed, at work
Province	character	Ontario, British Columbia, British Columbia, Ontario, Quebec, Ontario
Census.metropolitan.area	integer	0, 0, 9, 0, 0, 3
Age.group	character	65-69, 60-64, 70 and over, 45-49, 35-39, 30-34
Sex	integer	2, 2, 1, 2, 2, 2
Highest.educational.attainment	character	Postsecondary certificate or diploma, Above bachelor's degree, 0 to 8 years, Bachelor's degree, Postsecondary certificate or diploma, Bachelor's degree
Single.or.multiple.jobholder	character	NA, Single jobholder, NA, Single jobholder, Single jobholder, Single jobholder
Class.of.worker..main.job.	character	NA, private sector employees, NA, public sector employees, private sector employees, private sector employees
Type.of.work..main.job.	character	NA, Part-time, NA, Full-time, Full-time, Full-time
Occupation.at.main.job	character	NA, Sales and service, NA, Management, Manufacturing and utilities, Business, finance and administration
Usual.hours.worked.wk.at.main.job	double	NA, 25, NA, 37.5, 40, 33.7
Actual.hours.worked.wk.at.main.job	double	NA, 10, NA, 30, 42, 27.7

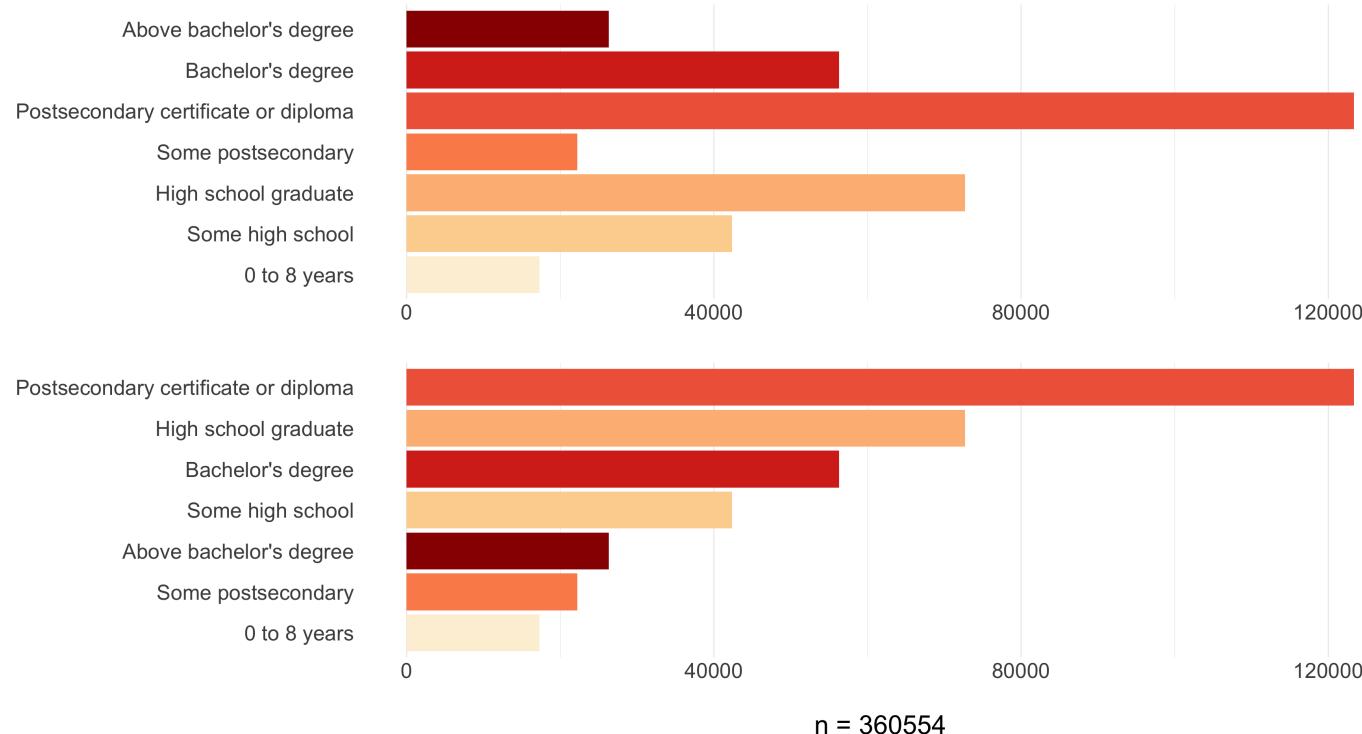
variable	class	first_values
Duration.of.unemployment	integer	NA, NA, NA, NA, NA, NA
Reason.for.part.time.work	character	NA, Personal preference, NA, NA, NA, NA
Reason.for.leaving.job	character	NA, NA, NA, NA, NA, NA
Usual.hourly.wages..employees.only.	double	NA, 15, NA, 53.33, 23, 36.52
Job.permanency..employees.only.	character	NA, Permanent, NA, Permanent, Permanent, Permanent
Flows.into.unemployment	character	NA, NA, NA, NA, NA, NA
Student.status	character	NA, Non-student, NA, Non-student, Non-student, Non-student
Statistical.Weight	integer	279, 235, 201, 217, 93, 696

Education	Count
0 to 8 years	17299
Above bachelor's degree	26337
Bachelor's degree	56276
High school graduate	72725
Postsecondary certificate or diploma	123358
Some high school	42341
Some postsecondary	22218

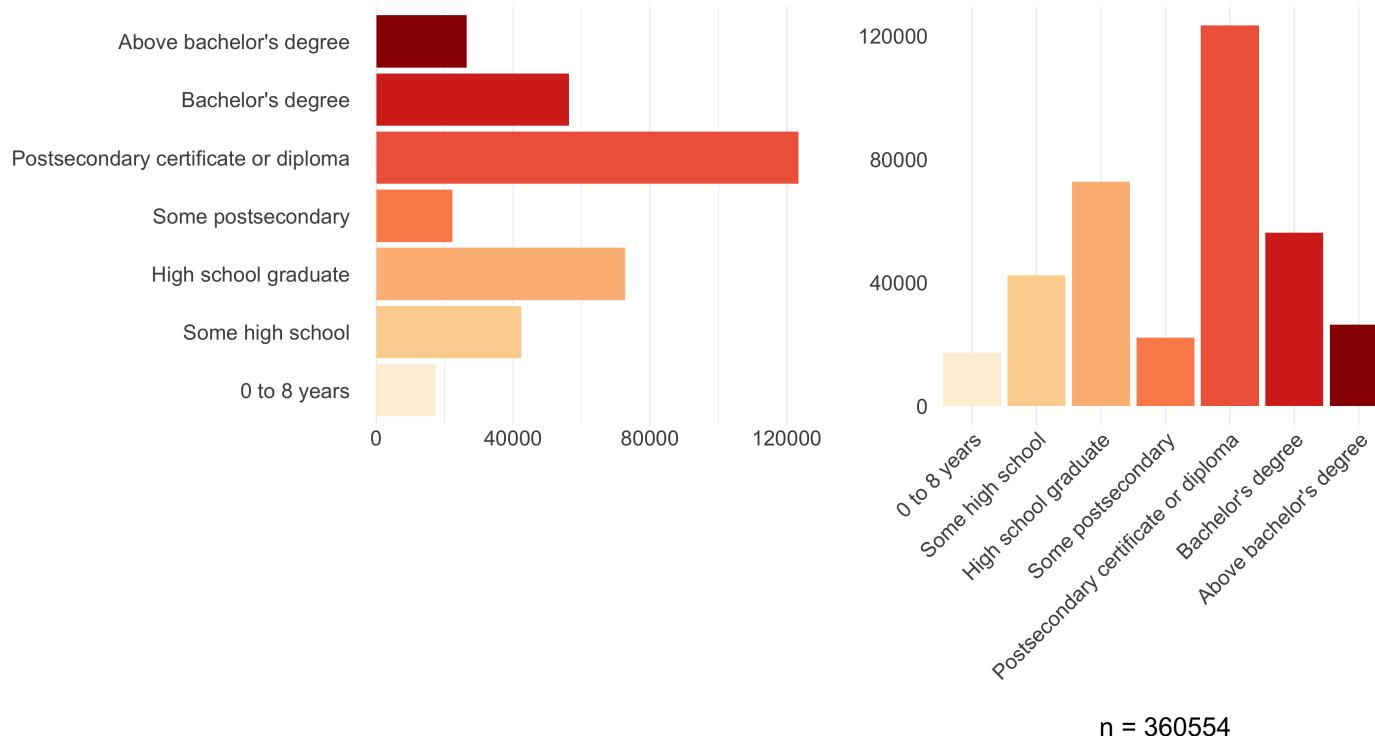
Labour Force Survey Data | Education



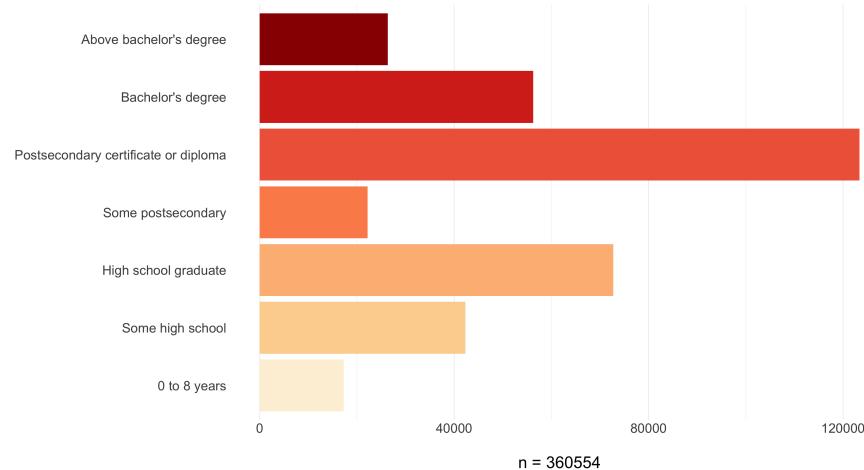
Labour Force Survey Data | Education



Labeling Options

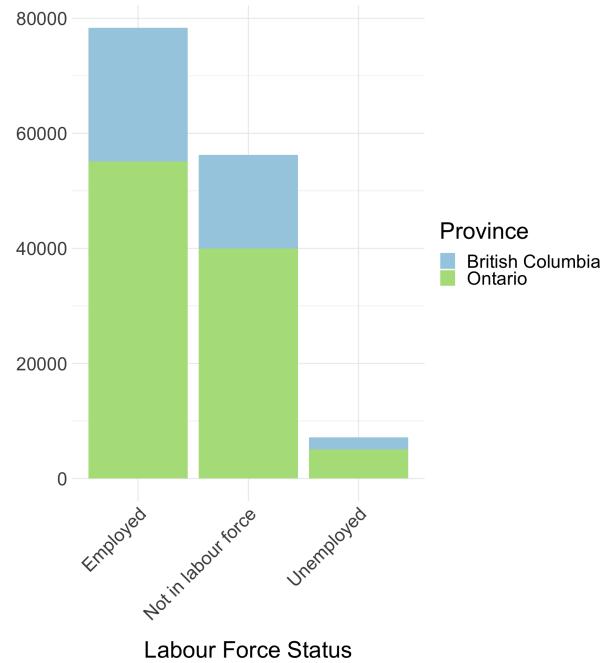
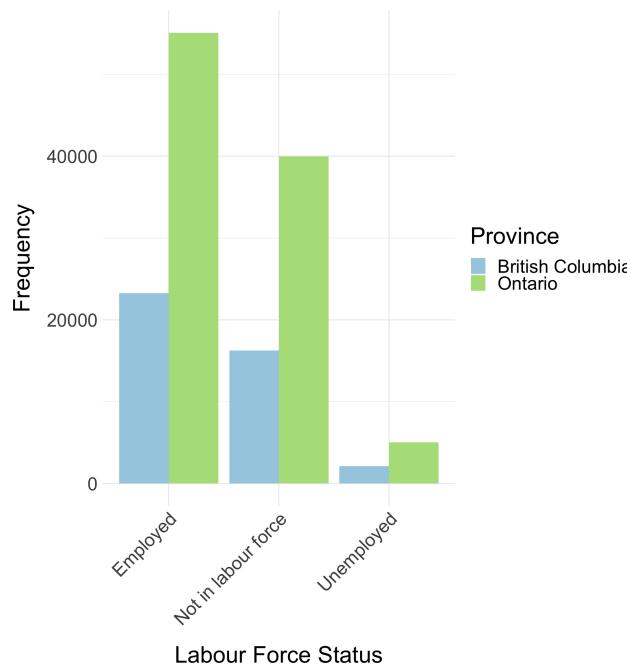


Bar Chart Design Considerations



- Graphs should start a baseline of 0.
- Bar widths should be equal for all categories.
- The bars should stand apart, not fused together.
- A total count should be provided in the legend.

Bar Chart Variations - Stacked & Side by Side



Numeric Data

Discrete = Counted

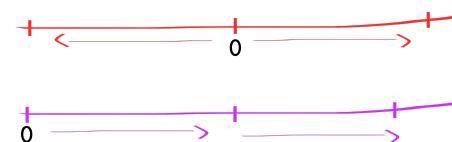
Continuous = Measured



Numeric Data

Interval = Greater or less than

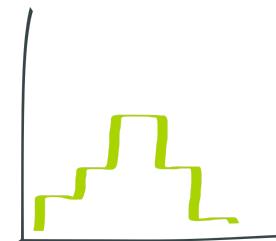
Ratio = Percentage more or less



Counts of Numeric Data

Visualization tool

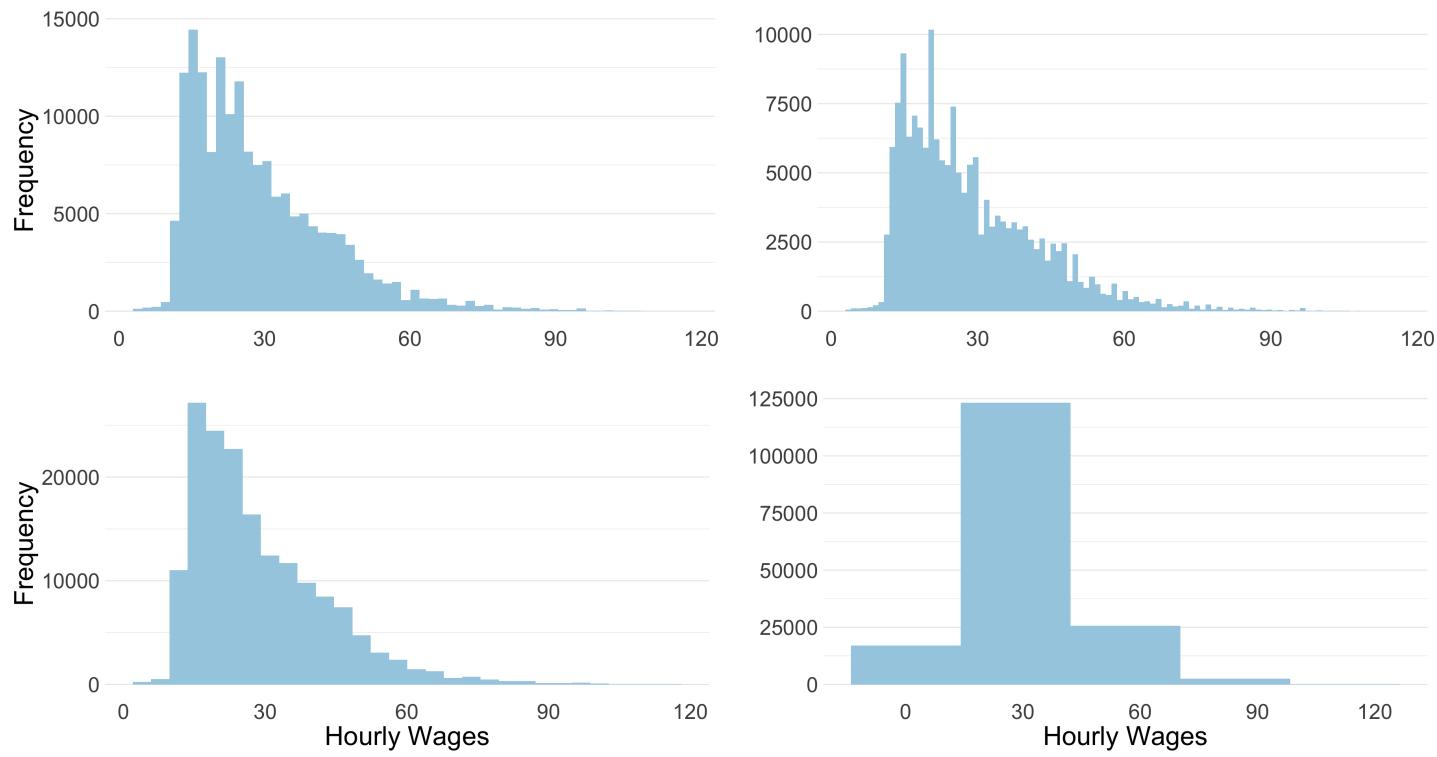
Frequency plots with histograms



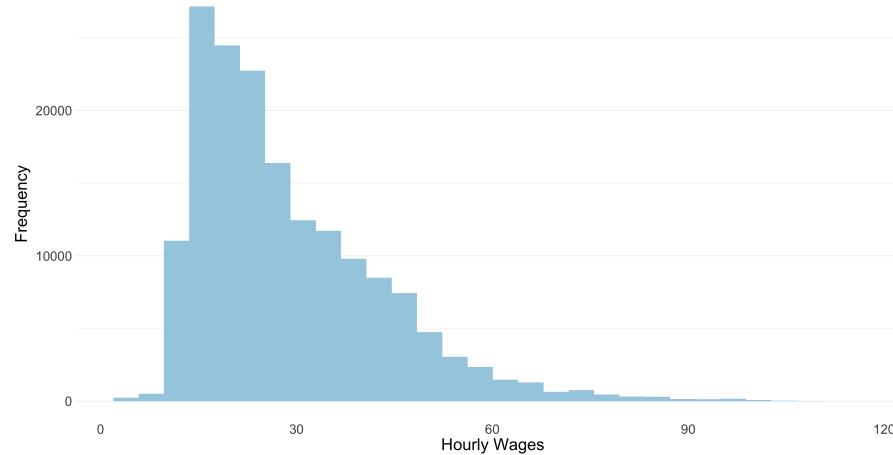
Labour Force survey Data | Wages



Binning Options



Histogram Design Considerations

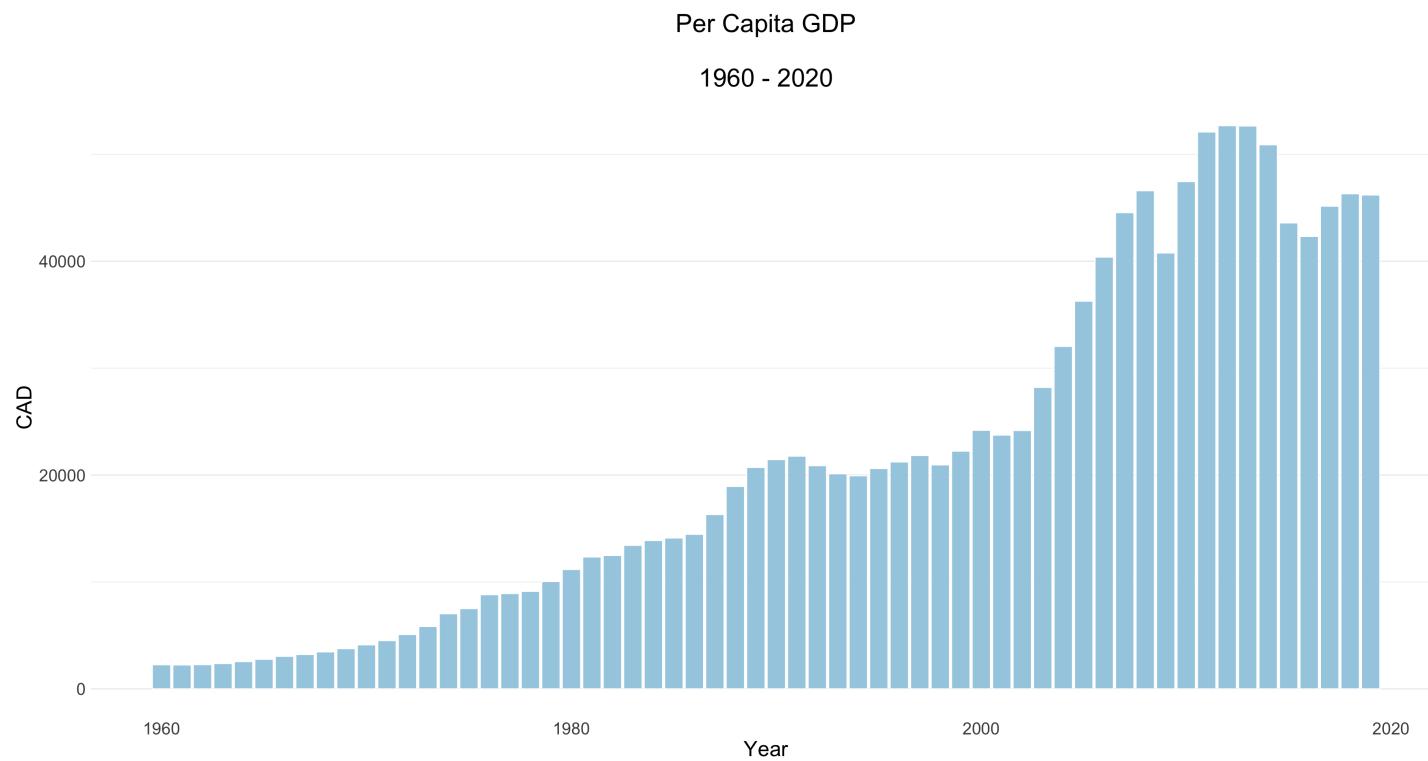


- Graphs should start a baseline of 0.
- The bars should be fused together.
- Bin width should be carefully considered.

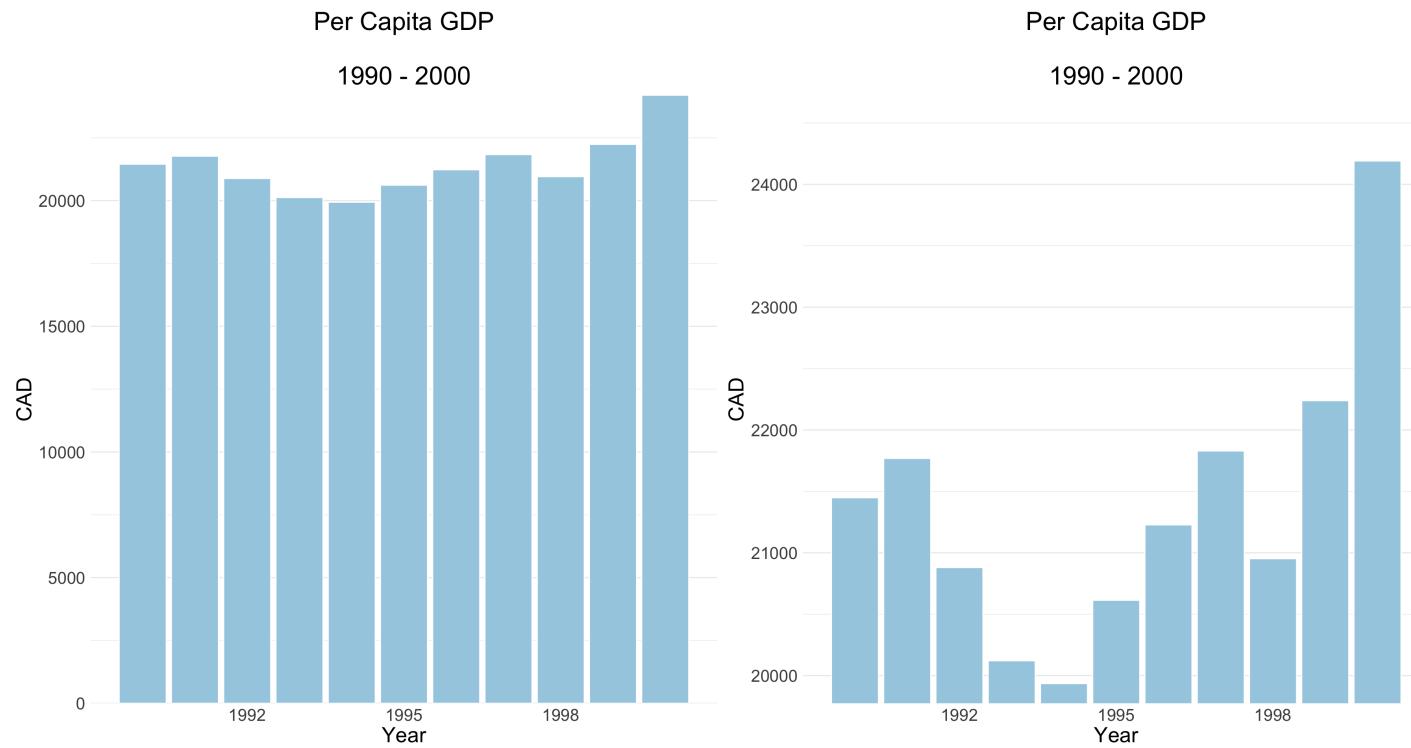
Year GDP.Per.Capita

2019	46194.73
2018	46313.17
2017	45148.55
2016	42322.48
2015	43585.51
2014	50893.45
2013	52652.59
2012	52678.39
2011	52087.45
2010	47448.01
2009	40773.06
2008	46594.45

Misleading Presentations



Misleading Presentations

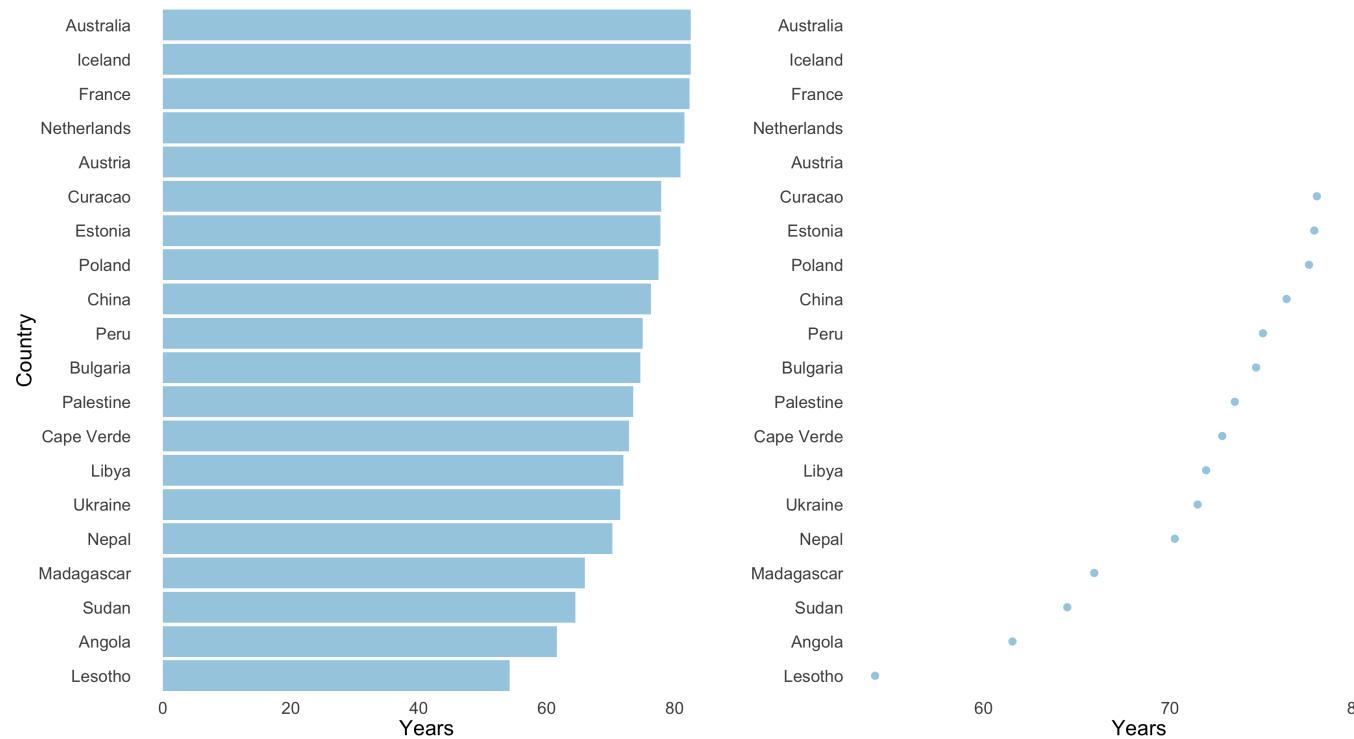


Dot Plots & Line Graphs

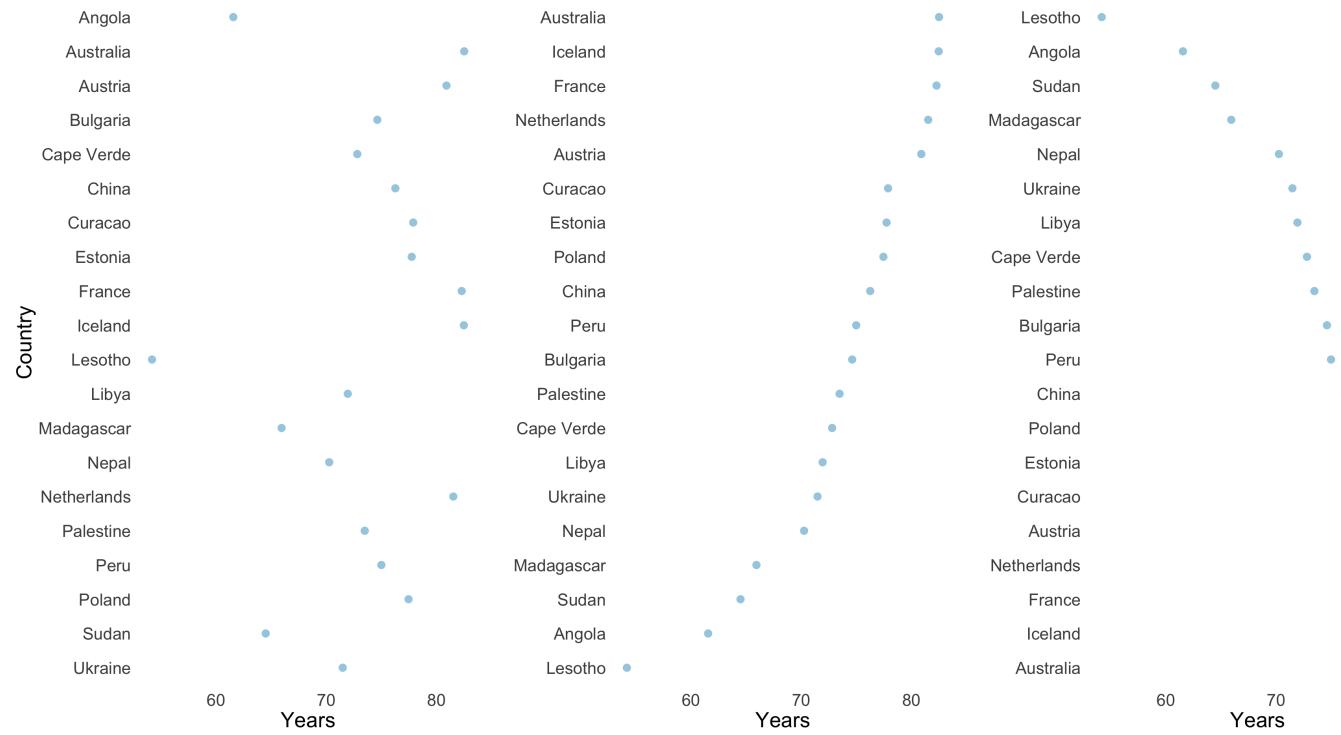
	Country	Code	Year	Life.expectency
7775	Peru	PER	2016	74.98300
2366	Curacao	CUW	2016	77.87317
1731	Cape Verde	CPV	2016	72.79800
5609	Lesotho	LSO	2016	54.17400
9941	Ukraine	UKR	2016	71.47634
9143	Sudan	SDN	2016	64.48600
456	Australia	AUS	2016	82.50000
228	Angola	AGO	2016	61.54700
7889	Poland	POL	2016	77.45122
1960	China	CHN	2016	76.25200
6889	Nepal	NPL	2016	70.25300
2998	Estonia	EST	2016	77.73659
4492	Iceland	ISL	2016	82.46829
513	Austria	AUT	2016	80.89024
3373	France	FRA	2016	82.27317
5723	Libya	LBY	2016	71.93400

	Country	Code	Year	Life.expectency
7547	Palestine	PSE	2016	73.47300
5974	Madagascar	MDG	2016	65.93200
1389	Bulgaria	BGR	2016	74.61463
6946	Netherlands	NLD	2016	81.50976

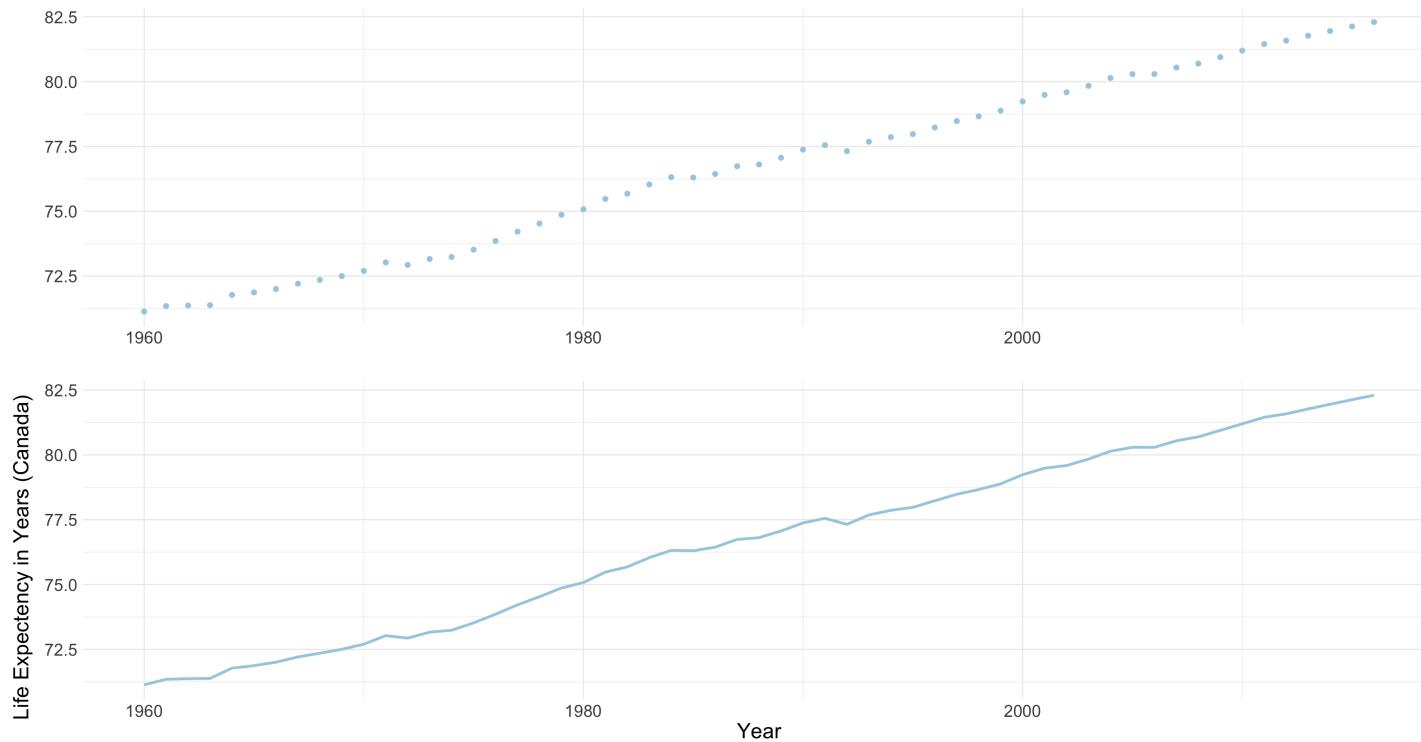
Volume vs Points



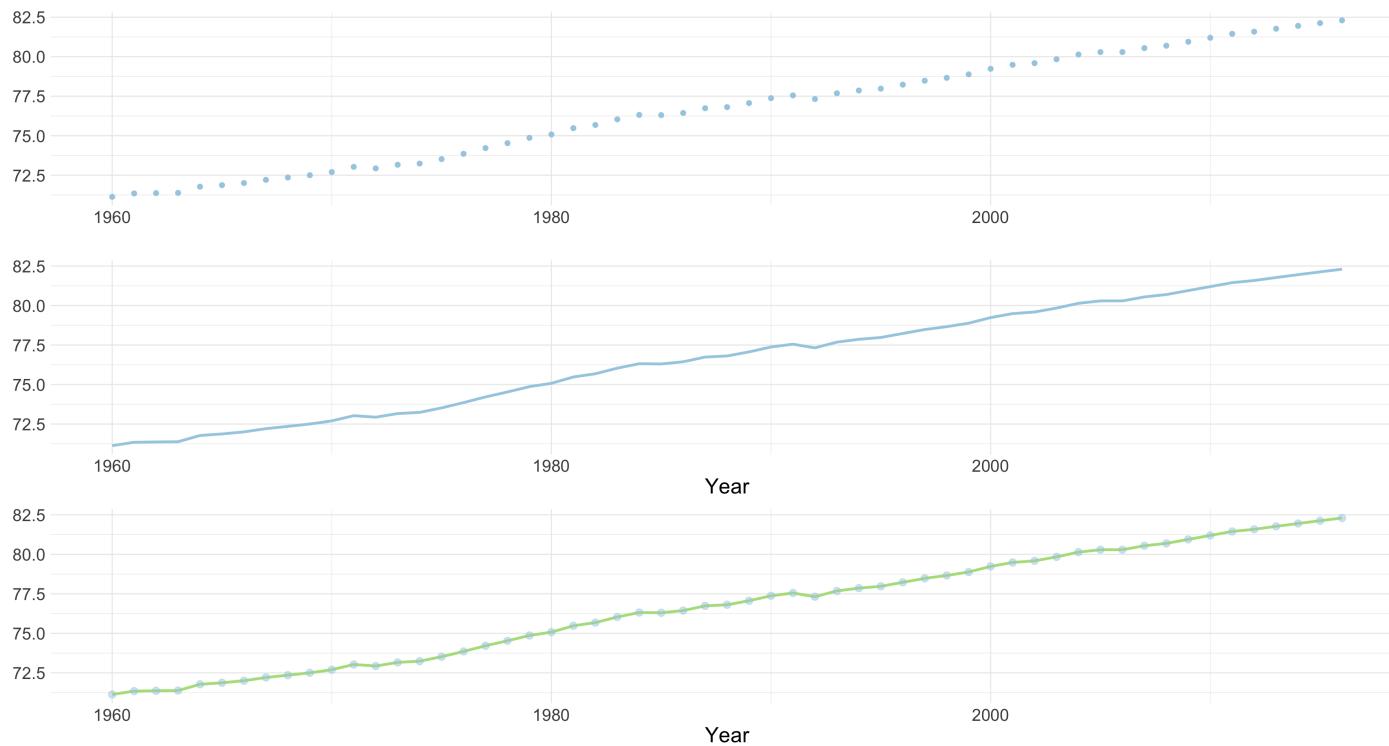
Visual Order



Time Sequences



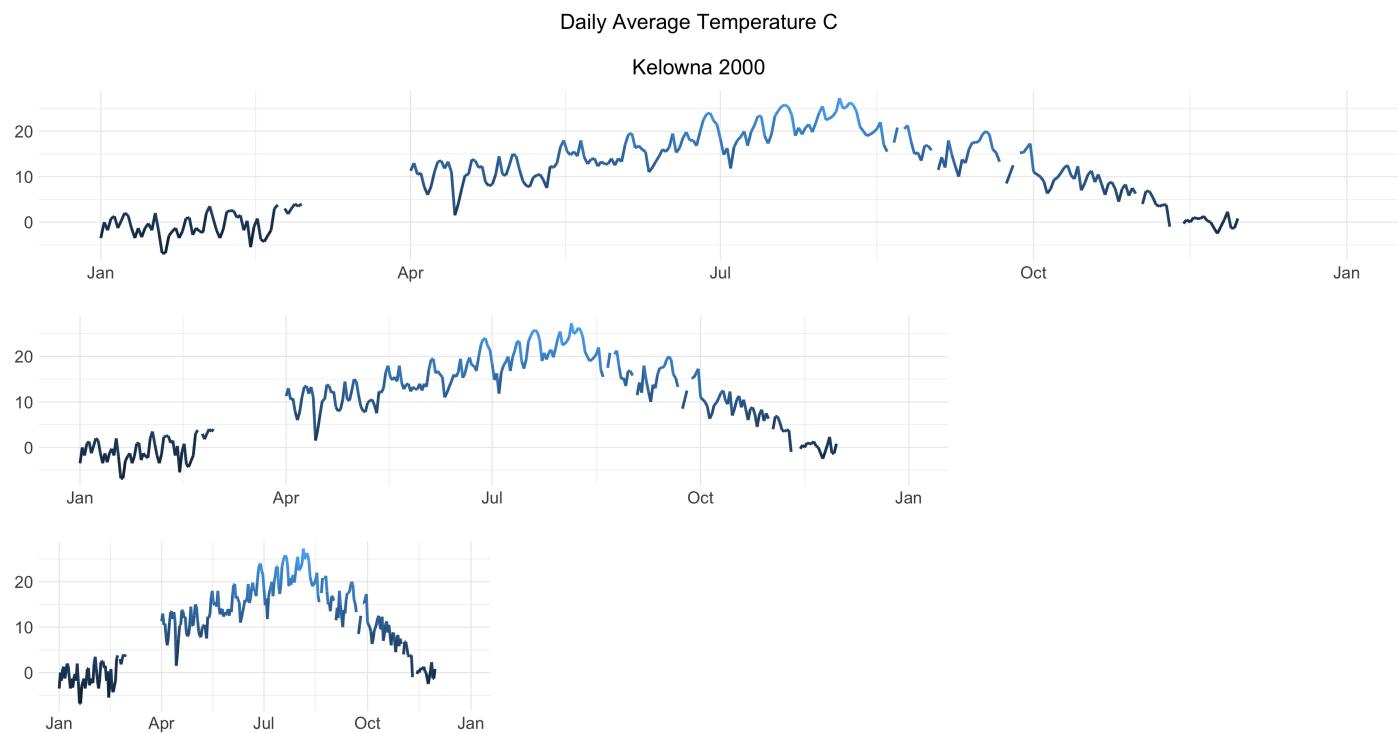
Dots & Lines



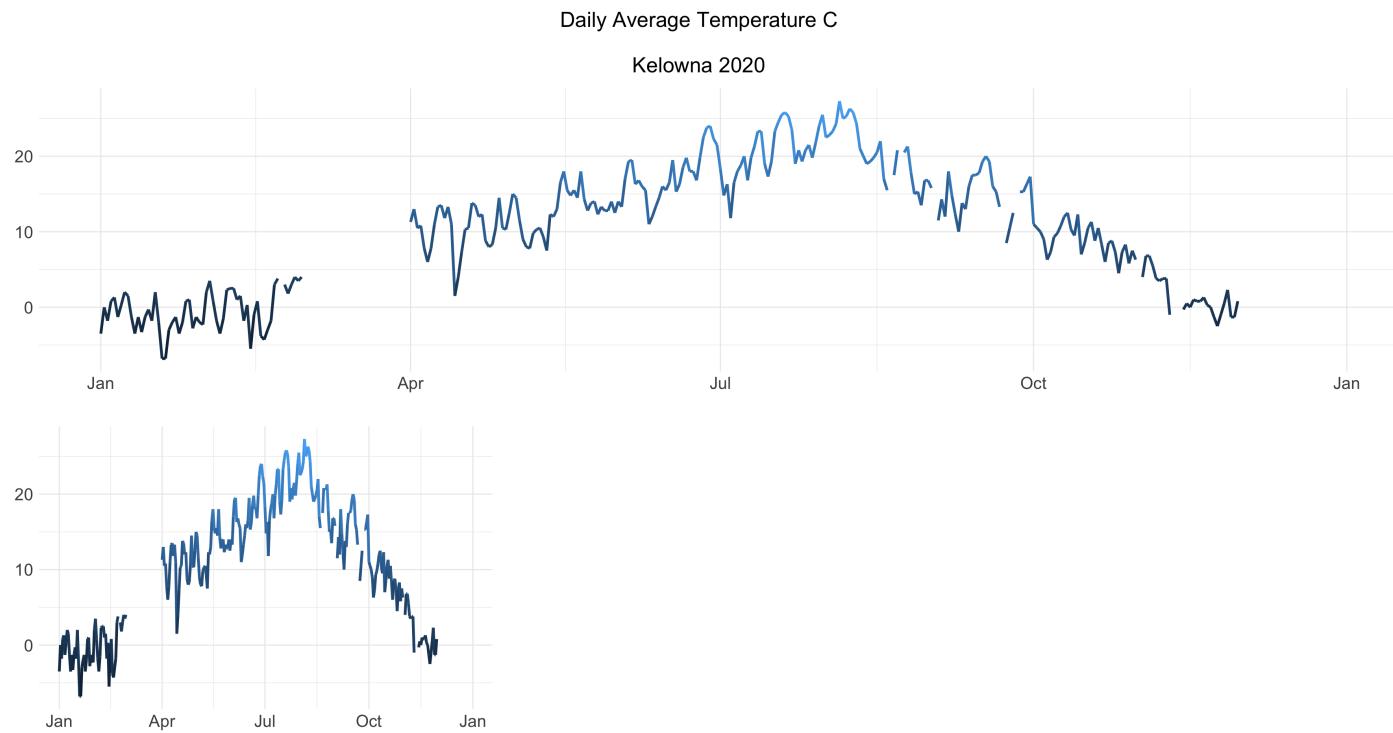
Longitude..x.	Latitude..y.	Station.Name	Climate.ID	Date.Time	Year	Month	Day	Data.Quality	Max.Temp..C.
-119.4	49.86	KELOWNA EAST	1123984	2000-01-01	2000	1	1	NA	-1.0
-119.4	49.86	KELOWNA EAST	1123984	2000-01-02	2000	1	2	NA	3.0
-119.4	49.86	KELOWNA EAST	1123984	2000-01-03	2000	1	3	NA	0.0
-119.4	49.86	KELOWNA EAST	1123984	2000-01-04	2000	1	4	NA	4.5
-119.4	49.86	KELOWNA EAST	1123984	2000-01-05	2000	1	5	NA	5.0
-119.4	49.86	KELOWNA EAST	1123984	2000-01-06	2000	1	6	NA	0.5
-119.4	49.86	KELOWNA EAST	1123984	2000-01-07	2000	1	7	NA	2.5
-119.4	49.86	KELOWNA EAST	1123984	2000-01-08	2000	1	8	NA	6.0
-119.4	49.86	KELOWNA EAST	1123984	2000-01-09	2000	1	9	NA	4.0
-119.4	49.86	KELOWNA EAST	1123984	2000-01-10	2000	1	10	NA	2.5

Longitude..x.	Latitude..y.	Station.Name	Climate.ID	Date.Time	Year	Month	Day	Data.Quality	Max.Temp..C.
-119.4	49.86	KELOWNA EAST	1123984	2000-01-11	2000	1	11	NA	0.5
-119.4	49.86	KELOWNA EAST	1123984	2000-01-12	2000	1	12	NA	3.0

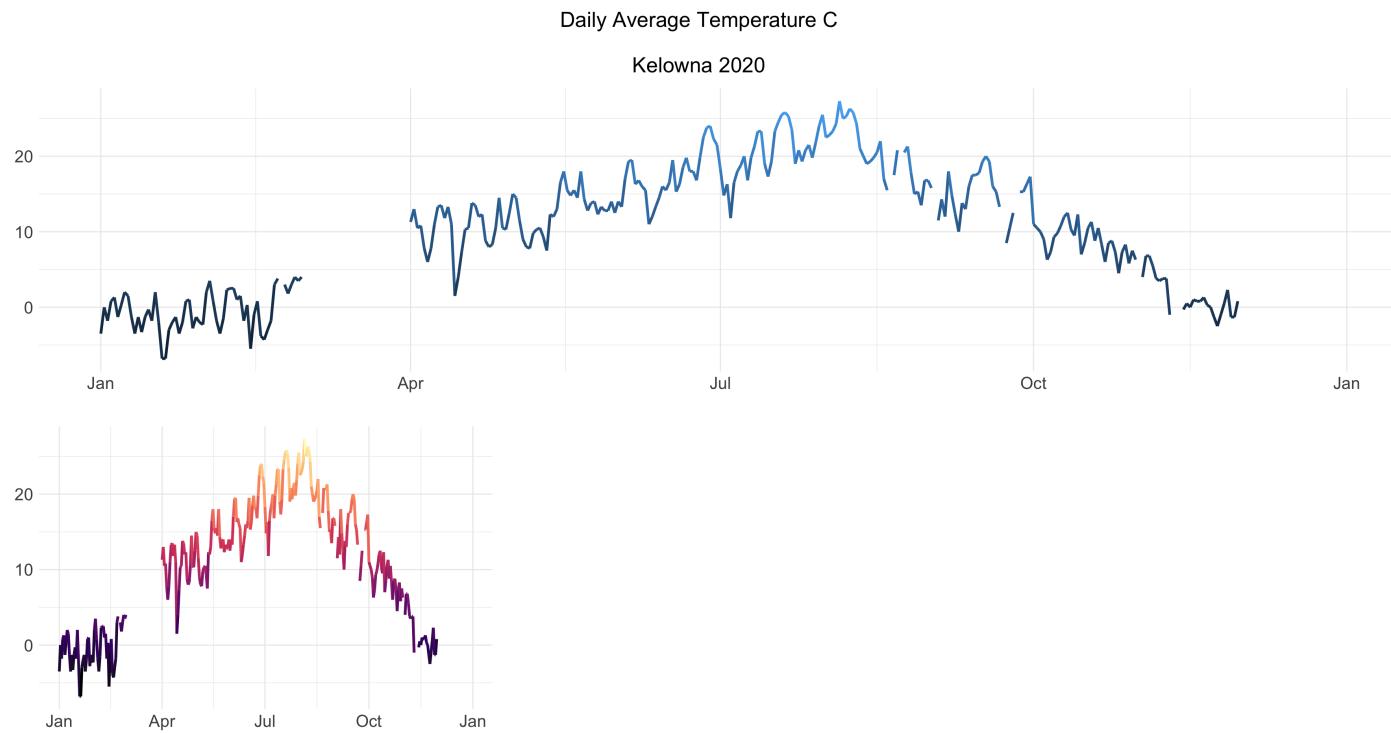
Misleading Presentations



Misleading Presentations



Misleading Presentations



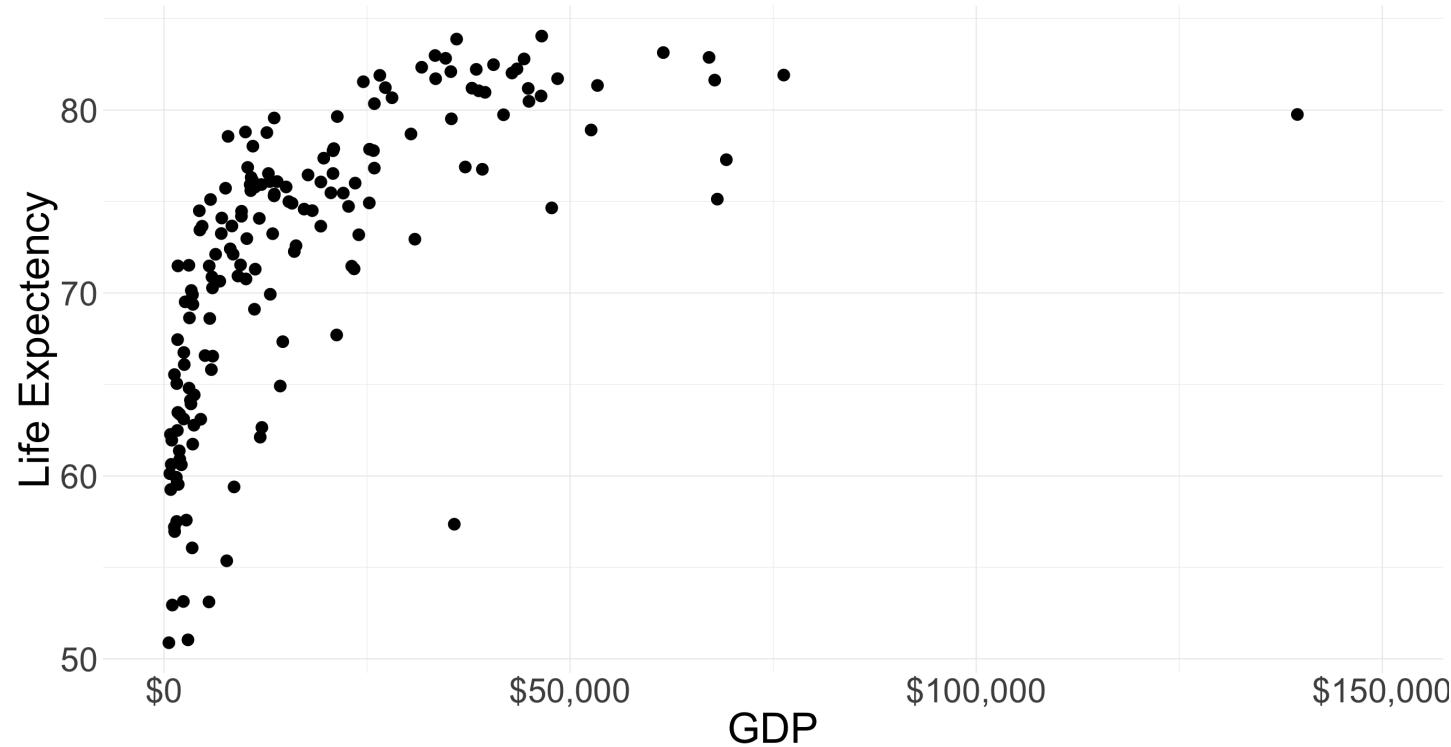
Layering Data and Statistics

Country	Code	Year	Population	Continent	Life.Expectency	GDP
Lesotho	LSO	2015	2059000	Africa	51.038	2954
Armenia	ARM	2015	2926000	Asia	74.467	9552
Uruguay	URY	2015	3412000	South America	77.369	19668
Slovakia	SVK	2015	5436000	Europe	76.827	25896
Bosnia and Herzegovina	BIH	2015	3429000	Europe	76.865	10305
Mali	MLI	2015	17439000	Africa	57.509	1563
Romania	ROU	2015	19925000	Europe	75.476	20549
Denmark	DNK	2015	5689000	Europe	80.475	44939
Jamaica	JAM	2015	2891000	North America	74.098	7115
Senegal	SEN	2015	14578000	Africa	66.747	2446
Eswatini	SWZ	2015	1104000	Africa	55.359	7726
Pakistan	PAK	2015	199427008	Asia	66.577	5056
Gambia	GMB	2015	2086000	Africa	60.910	1948
Haiti	HTI	2015	10696000	North America	62.485	1649
Vietnam	VNM	2015	92677000	Asia	75.110	5733
Congo	COG	2015	4856000	Africa	63.097	4526

Country	Code	Year	Population	Continent	Life.Expectency	GDP
Gabon	GAB	2015	1948000	Africa	64.913	14315
Nepal	NPL	2015	27015000	Asia	69.515	2607
North Macedonia	MKD	2015	2079000	Europe	75.406	13586
Cyprus	CYP	2015	1161000	Europe	80.350	25903

Country	Code	Year	Population	Continent	Life.Expectency	GDP
Lesotho	LSO	2015	2059000	Africa	51.038	2954
Armenia	ARM	2015	2926000	Asia	74.467	9552
Uruguay	URY	2015	3412000	South America	77.369	19668
Slovakia	SVK	2015	5436000	Europe	76.827	25896
Bosnia and Herzegovina	BIH	2015	3429000	Europe	76.865	10305
Mali	MLI	2015	17439000	Africa	57.509	1563
Romania	ROU	2015	19925000	Europe	75.476	20549
Denmark	DNK	2015	5689000	Europe	80.475	44939
Jamaica	JAM	2015	2891000	North America	74.098	7115
Senegal	SEN	2015	14578000	Africa	66.747	2446
Eswatini	SWZ	2015	1104000	Africa	55.359	7726
Pakistan	PAK	2015	199427008	Asia	66.577	5056
Gambia	GMB	2015	2086000	Africa	60.910	1948
Haiti	HTI	2015	10696000	North America	62.485	1649
Vietnam	VNM	2015	92677000	Asia	75.110	5733
Congo	COG	2015	4856000	Africa	63.097	4526

Country	Code	Year	Population	Continent	Life.Expectency	GDP
Gabon	GAB	2015	1948000	Africa	64.913	14315
Nepal	NPL	2015	27015000	Asia	69.515	2607
North Macedonia	MKD	2015	2079000	Europe	75.406	13586
Cyprus	CYP	2015	1161000	Europe	80.350	25903

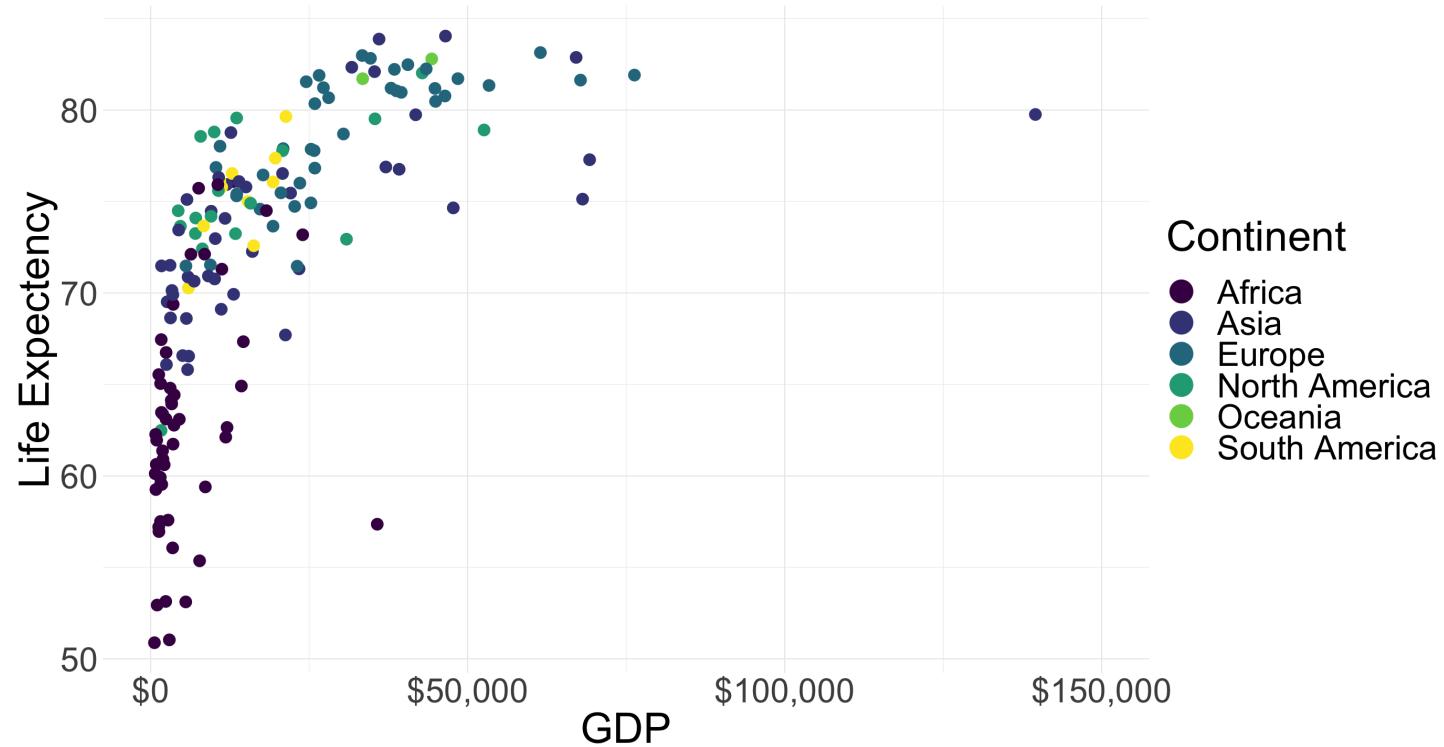


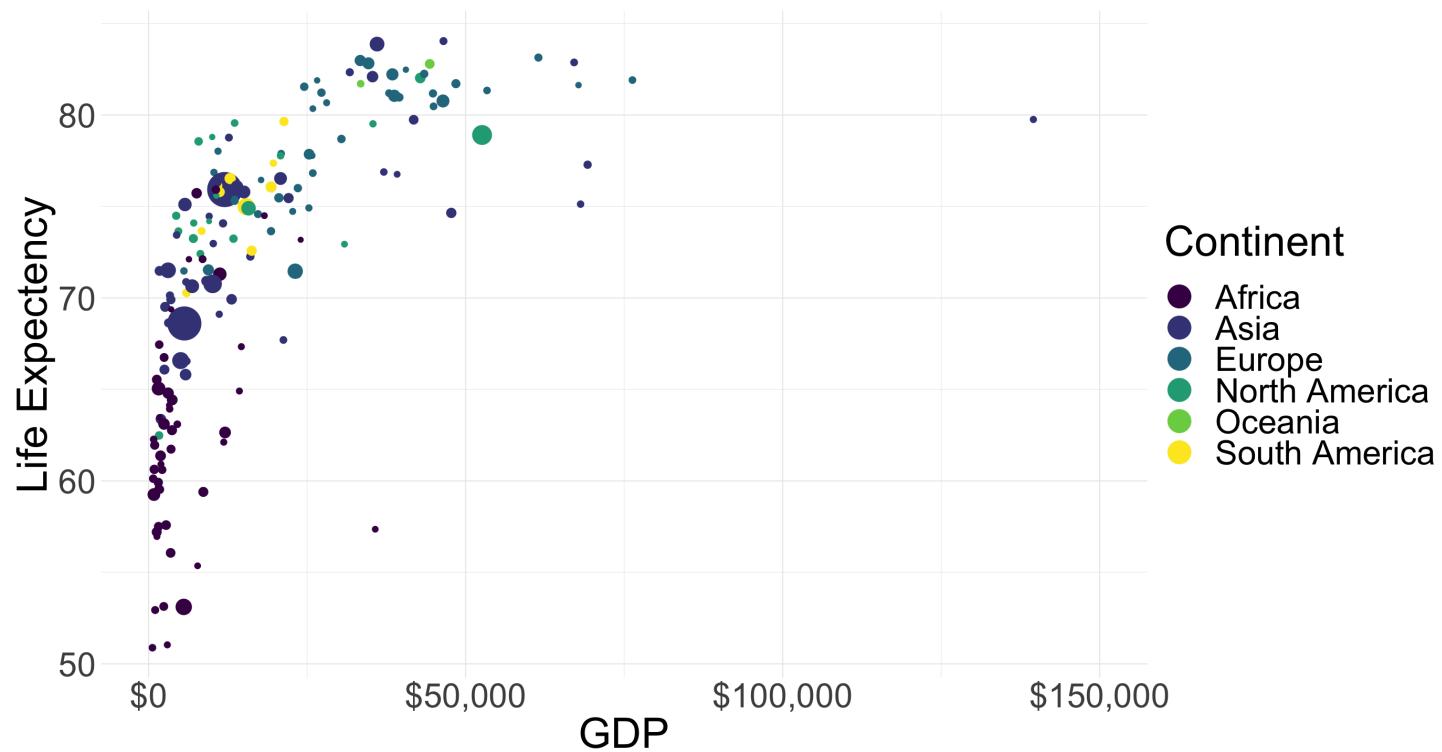
Country	Code	Year	Population	Continent	Life.Expectency	GDP
Lesotho	LSO	2015	2059000	Africa	51.038	2954
Armenia	ARM	2015	2926000	Asia	74.467	9552
Uruguay	URY	2015	3412000	South America	77.369	19668
Slovakia	SVK	2015	5436000	Europe	76.827	25896
Bosnia and Herzegovina	BIH	2015	3429000	Europe	76.865	10305
Mali	MLI	2015	17439000	Africa	57.509	1563
Romania	ROU	2015	19925000	Europe	75.476	20549
Denmark	DNK	2015	5689000	Europe	80.475	44939
Jamaica	JAM	2015	2891000	North America	74.098	7115
Senegal	SEN	2015	14578000	Africa	66.747	2446
Eswatini	SWZ	2015	1104000	Africa	55.359	7726
Pakistan	PAK	2015	199427008	Asia	66.577	5056
Gambia	GMB	2015	2086000	Africa	60.910	1948
Haiti	HTI	2015	10696000	North America	62.485	1649
Vietnam	VNM	2015	92677000	Asia	75.110	5733
Congo	COG	2015	4856000	Africa	63.097	4526

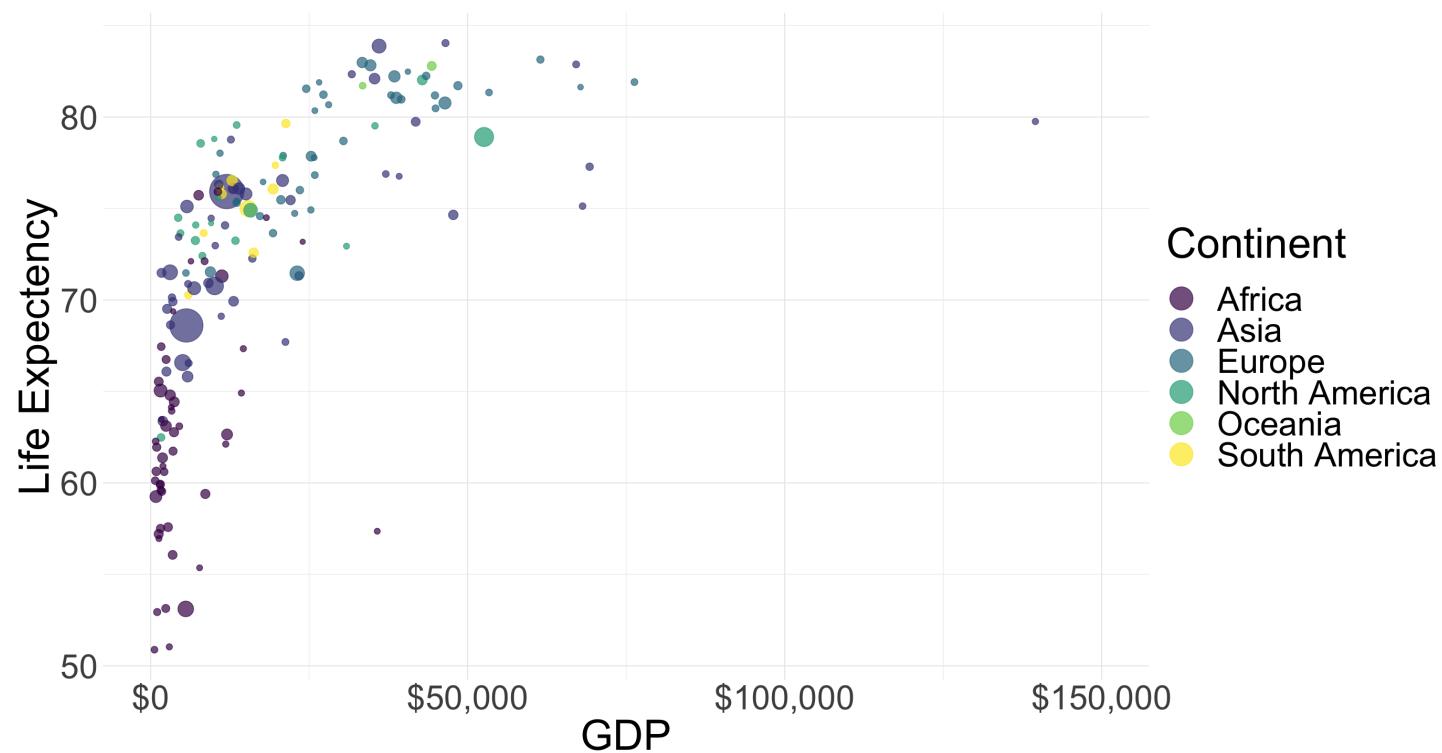
Country	Code	Year	Population	Continent	Life.Expectency	GDP
Gabon	GAB	2015	1948000	Africa	64.913	14315
Nepal	NPL	2015	27015000	Asia	69.515	2607
North Macedonia	MKD	2015	2079000	Europe	75.406	13586
Cyprus	CYP	2015	1161000	Europe	80.350	25903

Country	Code	Year	Population	Continent	Life.Expectency	GDP
Lesotho	LSO	2015	2059000	Africa	51.038	2954
Armenia	ARM	2015	2926000	Asia	74.467	9552
Uruguay	URY	2015	3412000	South America	77.369	19668
Slovakia	SVK	2015	5436000	Europe	76.827	25896
Bosnia and Herzegovina	BIH	2015	3429000	Europe	76.865	10305
Mali	MLI	2015	17439000	Africa	57.509	1563
Romania	ROU	2015	19925000	Europe	75.476	20549
Denmark	DNK	2015	5689000	Europe	80.475	44939
Jamaica	JAM	2015	2891000	North America	74.098	7115
Senegal	SEN	2015	14578000	Africa	66.747	2446
Eswatini	SWZ	2015	1104000	Africa	55.359	7726
Pakistan	PAK	2015	199427008	Asia	66.577	5056
Gambia	GMB	2015	2086000	Africa	60.910	1948
Haiti	HTI	2015	10696000	North America	62.485	1649
Vietnam	VNM	2015	92677000	Asia	75.110	5733
Congo	COG	2015	4856000	Africa	63.097	4526

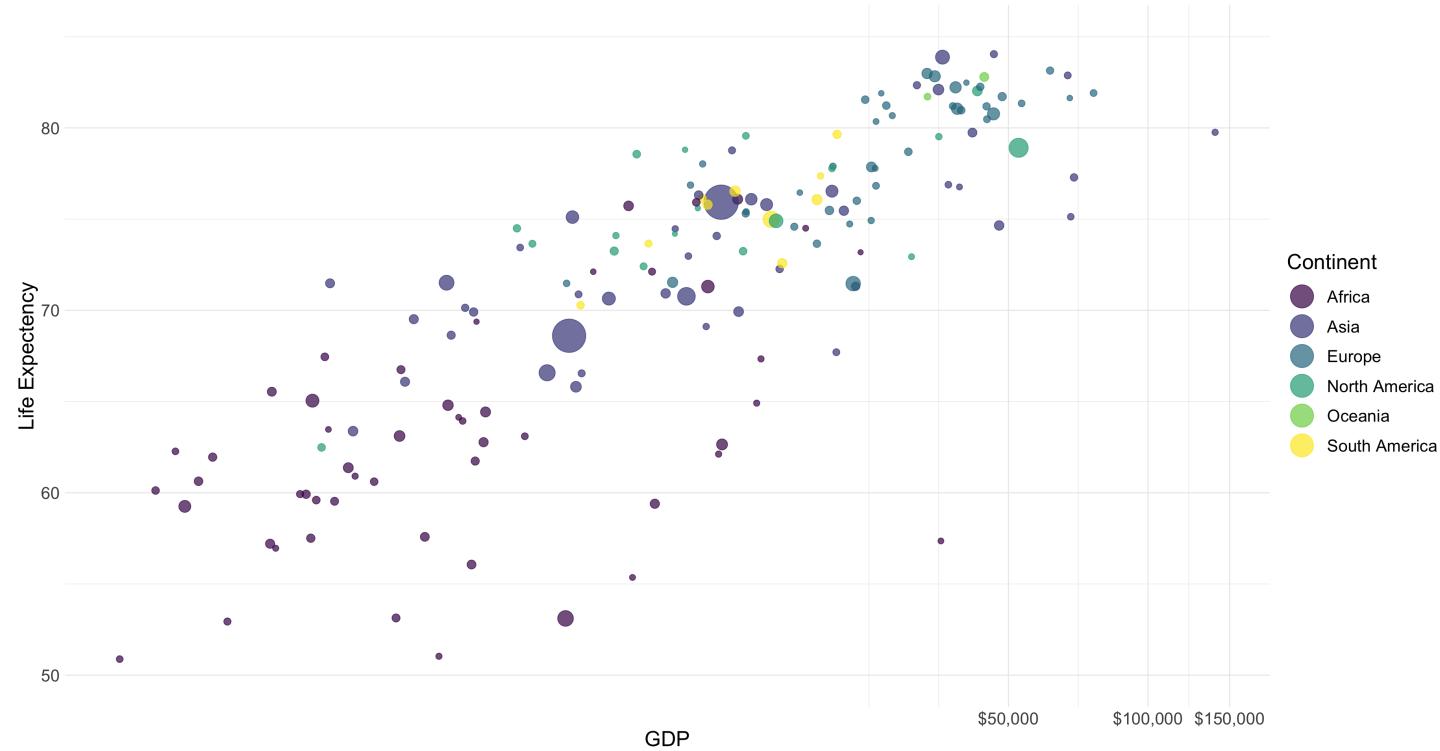
Country	Code	Year	Population	Continent	Life.Expectency	GDP
Gabon	GAB	2015	1948000	Africa	64.913	14315
Nepal	NPL	2015	27015000	Asia	69.515	2607
North Macedonia	MKD	2015	2079000	Europe	75.406	13586
Cyprus	CYP	2015	1161000	Europe	80.350	25903

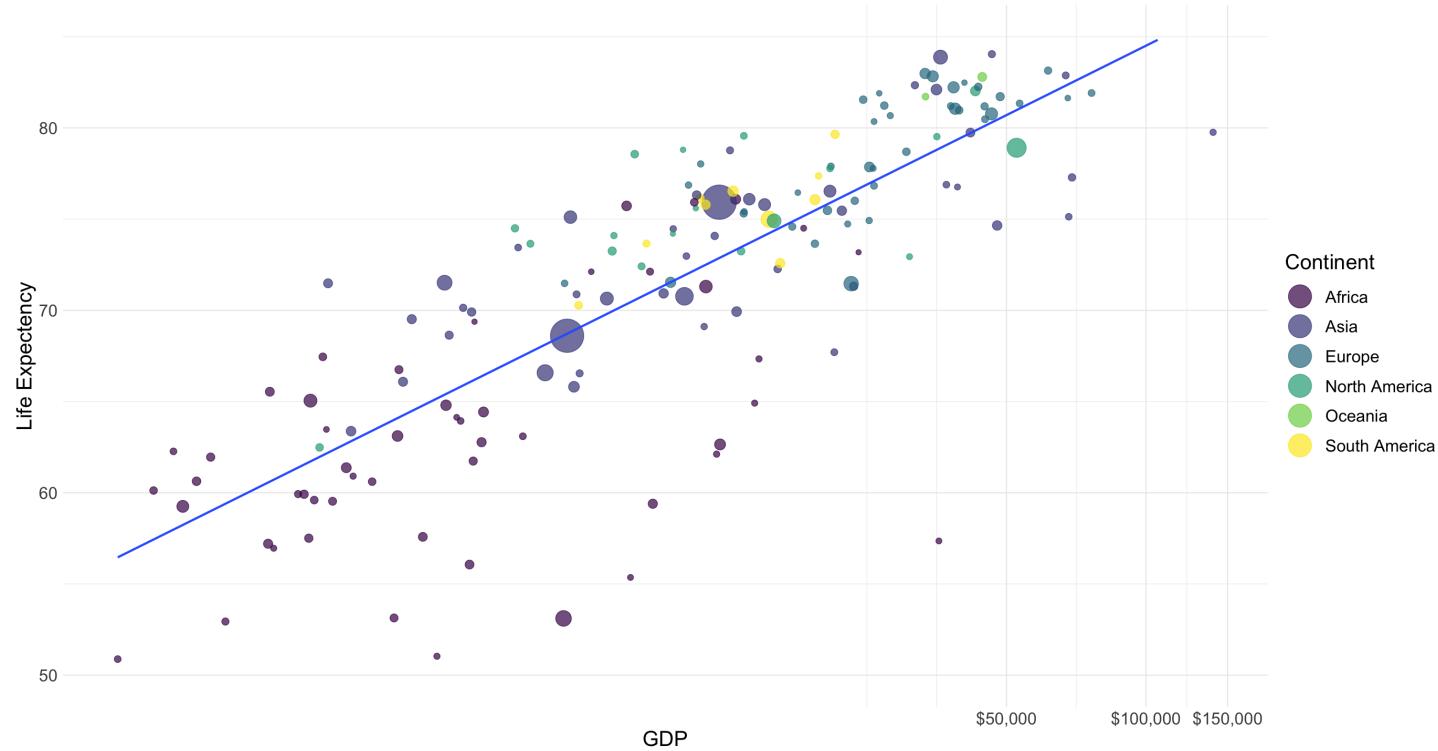


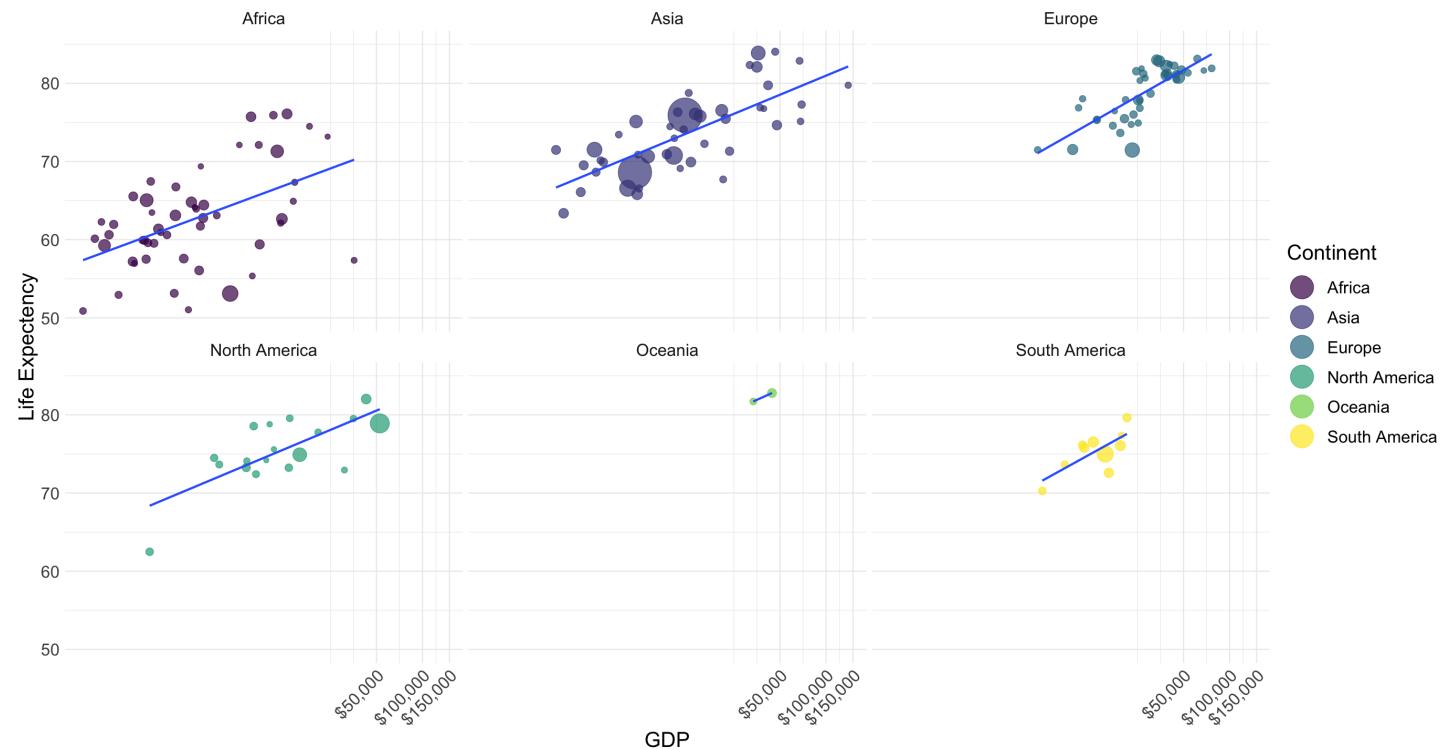




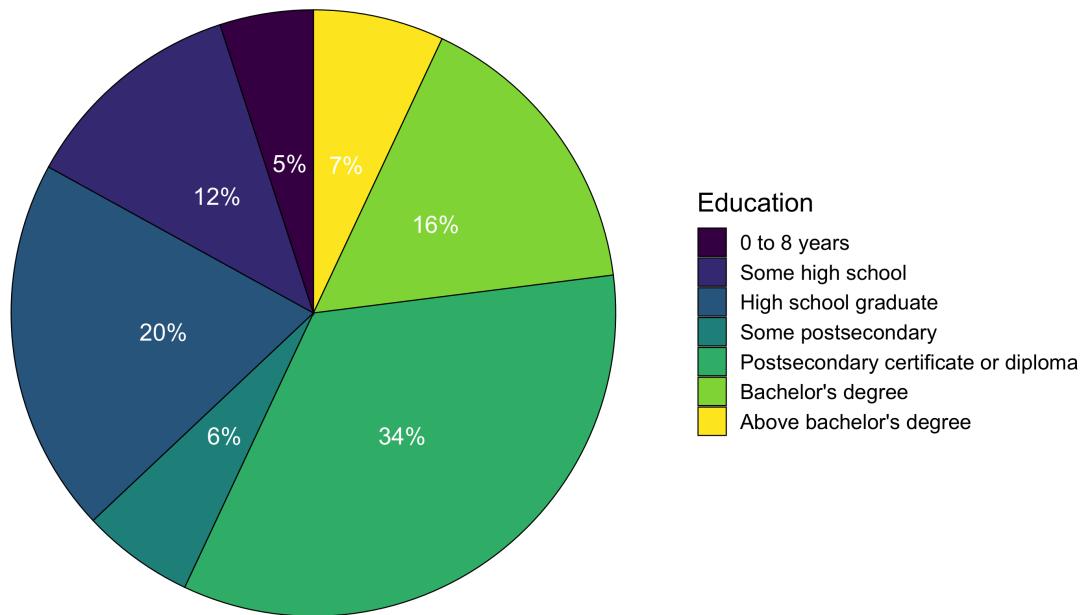


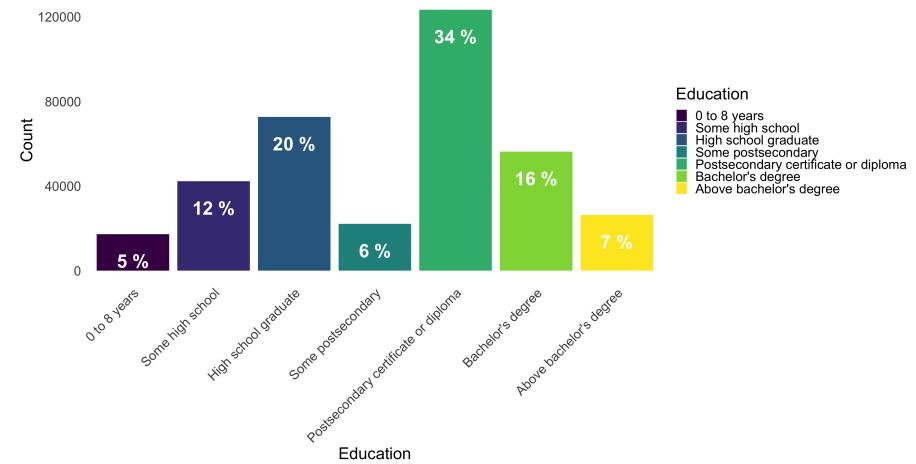
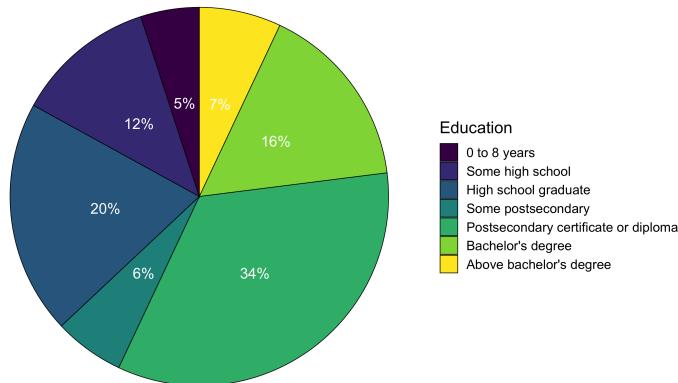






Pie Charts





Finding Data

CBC Fatal Police Encounters



Tracking (In)Justice

The CBC data has been critiqued for lacking a transparent methodology. Additionally, policing scholars have noted that while certain practices in the collection of data by the CBC data meet journalistic standards, they may not meet scholarly research standards.

Guiding Questions

Who would be responsible for

- Generating or producing the data
- Keeping a record of the data
- Reporting out on the data

Who would have an interest in keeping a record of the data

Data Aggregators

- Governments
 - Municipal, Provincial/Sate, Federal/National
 - [City of Kelowna open data portal](#)
- UN Agencies
 - World Health Organization (WHO) [Data](#)
- Intergovernmental Organizations
 - Organization for Economic Co-Operation and Development (OECD) [Data](#)
- International Bodies
 - [World Bank Data](#)

UBC Data Guide

- Need more? A list of public access data sources on a range of topics
 - [https://guides.library.ubc.ca/c.php?
g=712037&p=5075056](https://guides.library.ubc.ca/c.php?g=712037&p=5075056)

UBC Purchased Data Sets

Purchased and leased data sets are available in a couple of ways:

- [Abacus](#), where we host data sets
- [On our databases page](#), where we link to data we don't host.

Statistics Canada

- Table data
 - aggregated
 - free
- Public Use Microdata Files
 - non aggregated, anonymized
 - fee based, but increasingly less so
 - available through UBC free of charge
- Raw survey data
 - highly restricted