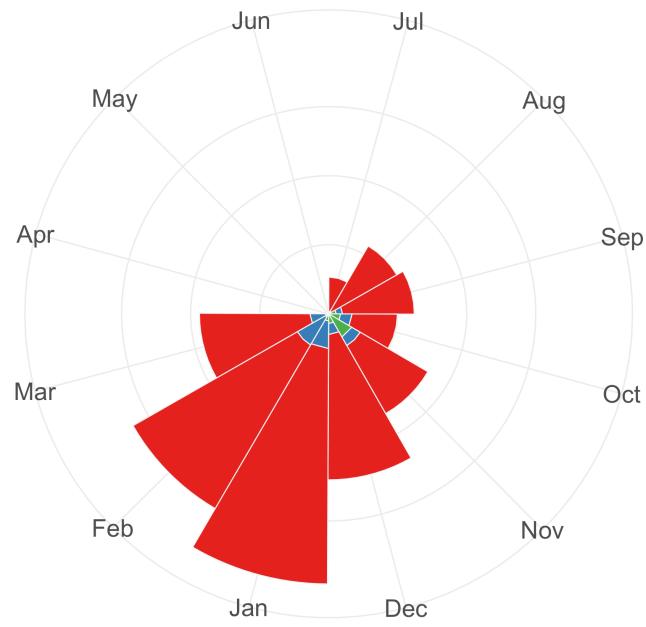


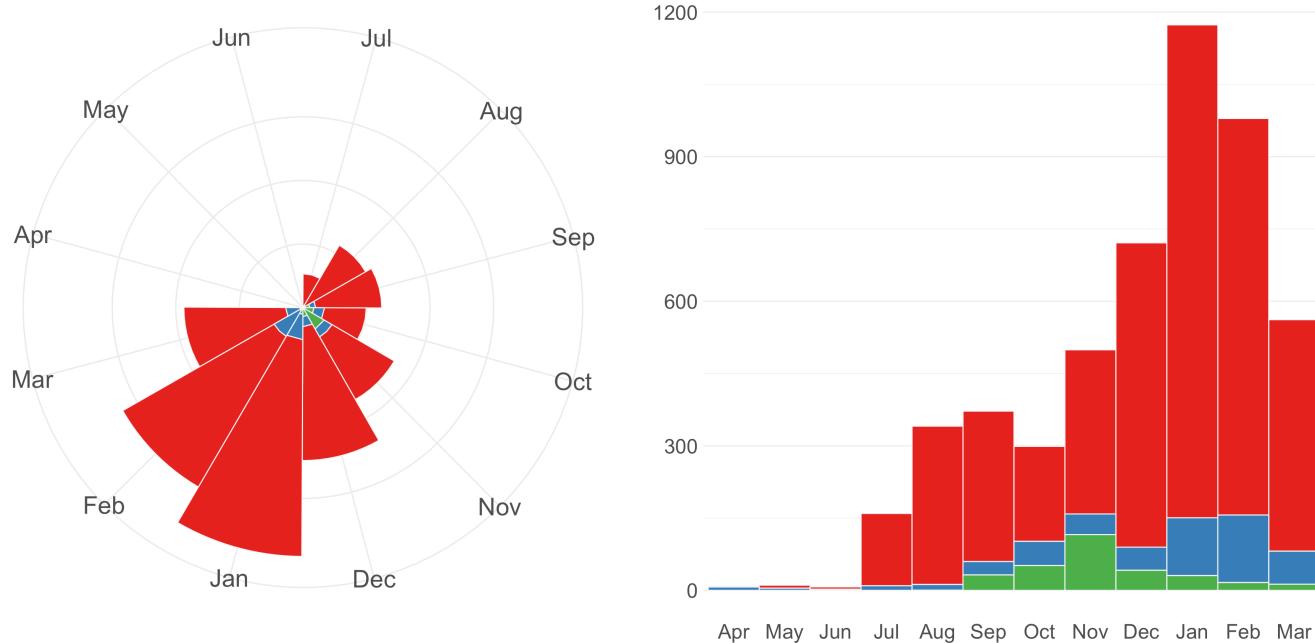
# **Data Acquisition & Visualization**

**CORH 205**

Mathew Vis-Dunbar

March 2022





# Overview

- Acquisition, Reliability and Trust
- Organization and Data Set Structure
- Data Types
- Data Processing or Interpretation
- Visual Layering of Content
- Visualizing Raw vs Statistical Computations of Data
- Tools for Visualizing Data

# Overview

- Acquisition, Reliability and Trust
- Organization and Data Set Structure
- Data Types
- Data Processing or Interpretation
- Visual Layering of Content
- Visualizing Raw vs Statistical Computations of Data
- Tools for Visualizing Data

# **What We're Not Talking About**

- 3D visualizations
- Info graphics

# **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.

# An Example

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

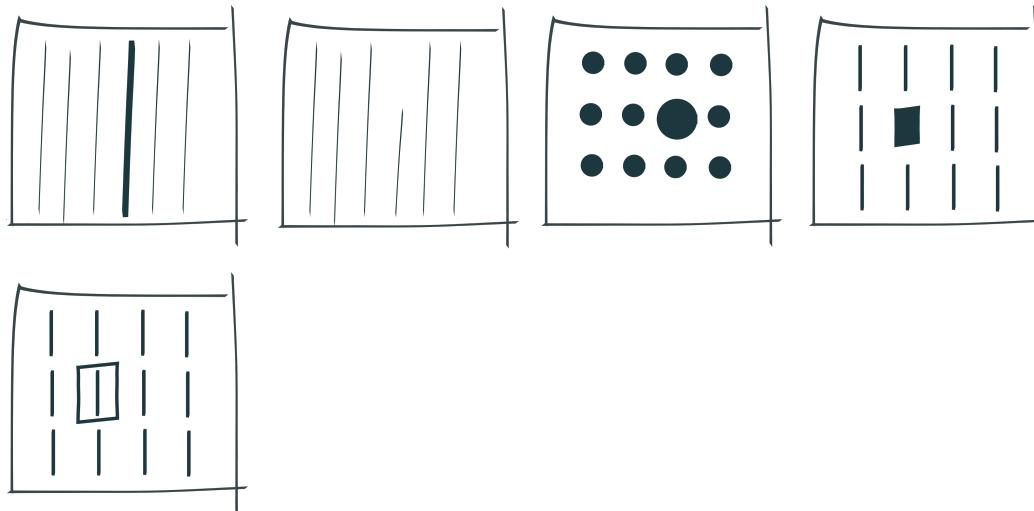
# An Example

3	9	7	9	9	1	9	<b>5</b>	3	<b>5</b>
3	9	<b>5</b>	2	4	2	7	<b>5</b>	8	6
2	9	7	<b>5</b>	6	4	3	8	3	8
6	3	9	8	8	<b>5</b>	8	3	8	4
3	8	9	2	6	6	9	2	1	7
4	7	7	1	6	3	3	2	7	4
6	9	<b>5</b>	9	7	9	7	6	7	3
9	3	7	9	1	4	3	4	7	9
2	4	<b>5</b>	6	6	6	7	1	6	7
3	1	6	<b>5</b>	2	9	6	6	7	6

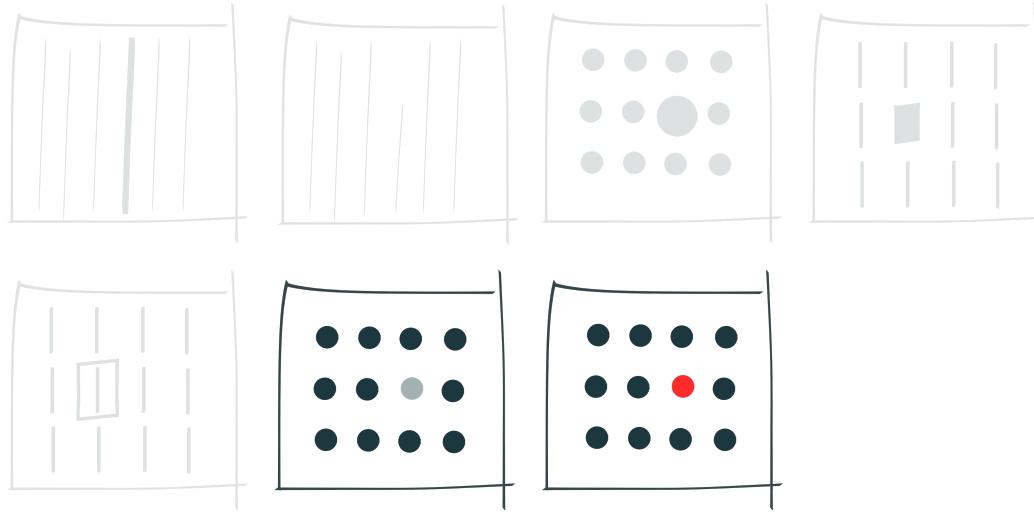
# **Preattentive Attributes**

- Form
- Colour
- Position

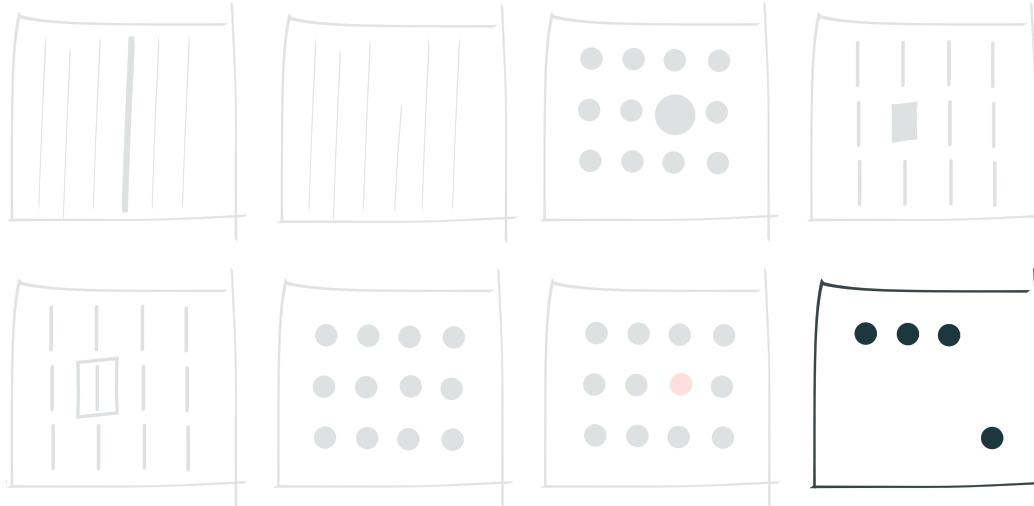
# Form



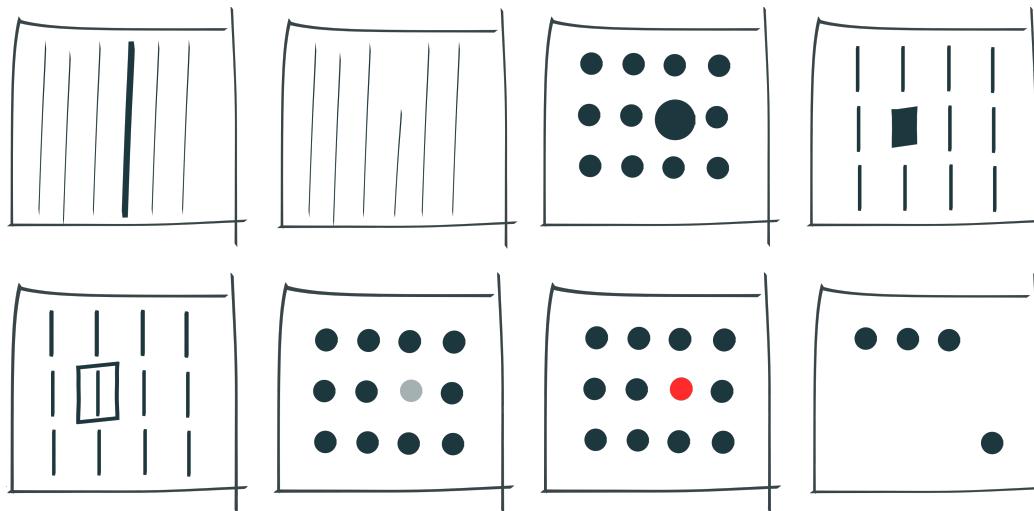
# Colour



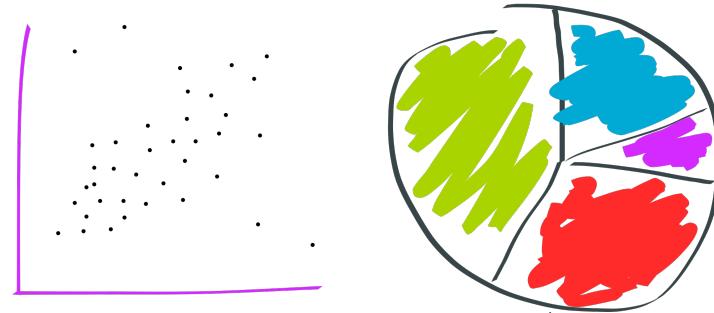
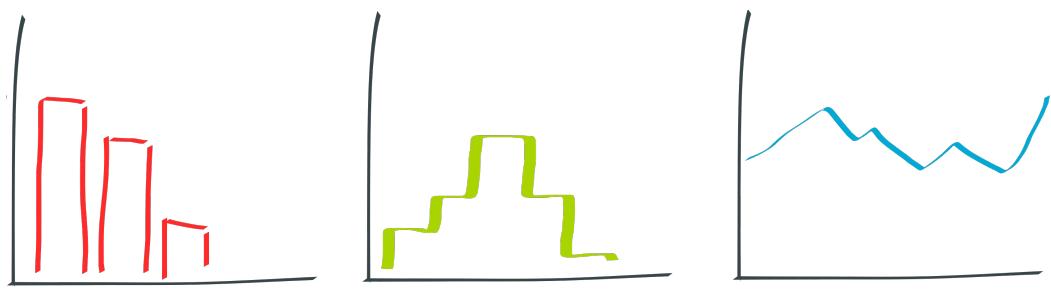
# Position

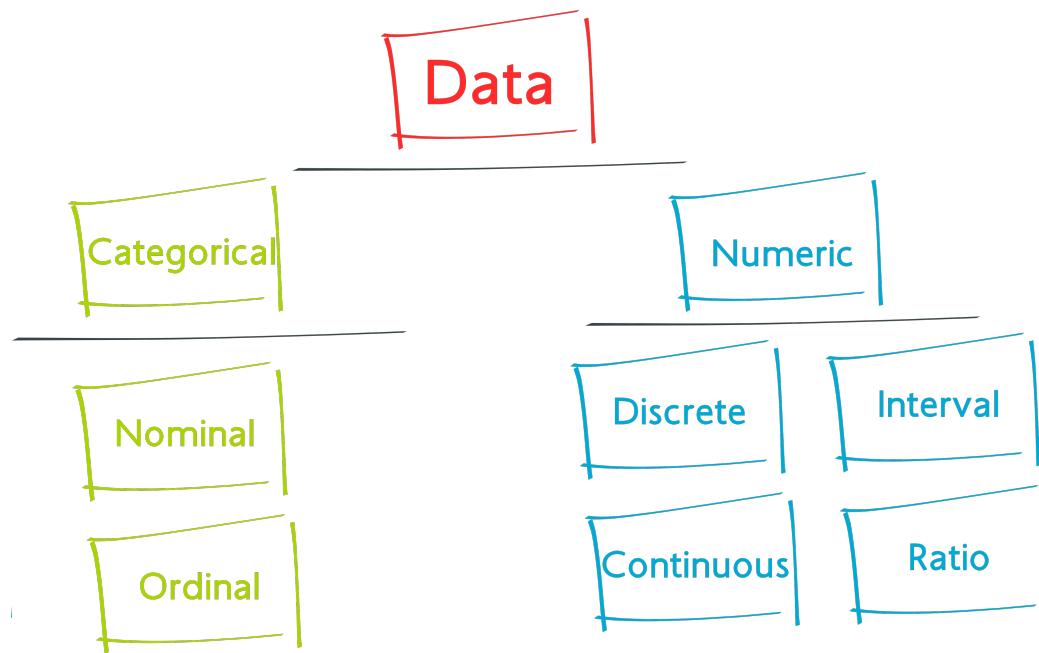


# Preattentive Attributes



## **Types of Visualisations**





# Categorical Data

## Nominal data

No order



## Ordinal data

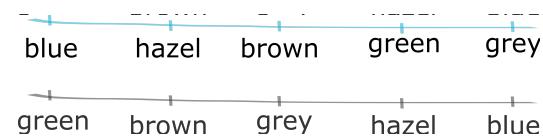
Intrinsic order



# Categorical Data

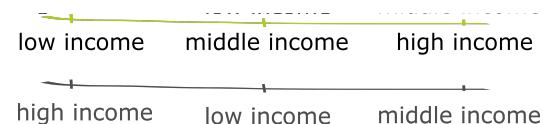
## Nominal data

No order



## Ordinal data

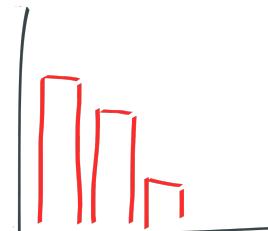
Intrinsic order



# **Bar Plots**

**Visualization tool**

**Frequency plots with bar charts**

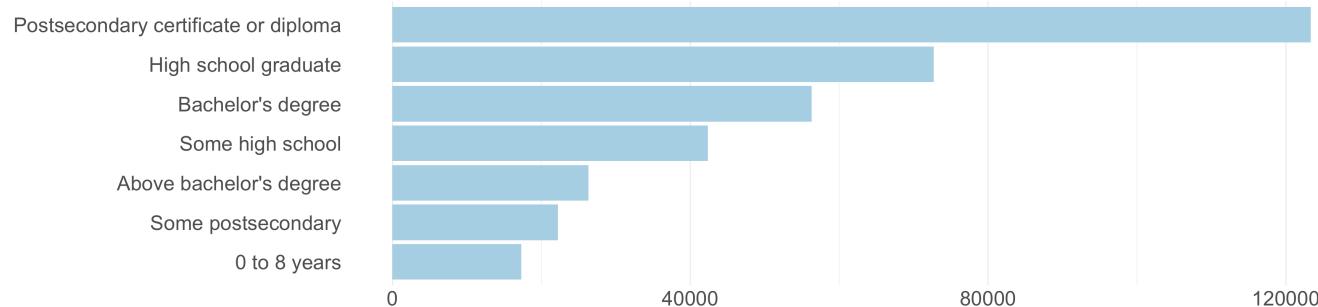
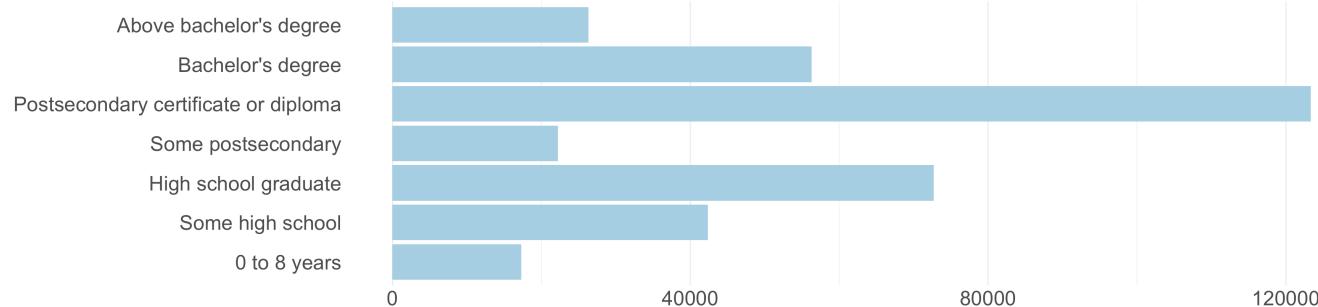


<b>variable</b>	<b>class</b>	<b>first_values</b>
Record.number	integer	1, 2, 3, 4, 5, 6
Survey.year	integer	2020, 2020, 2020, 2020, 2020, 2020
Survey.month	character	January, January, January, January, January, January
Labour.force.status	character	Not in labour force, Employed, at work, Not in labour force, Employed, at work, Employed, at work, 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

<b>variable</b>	<b>class</b>	<b>first_values</b>
		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
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

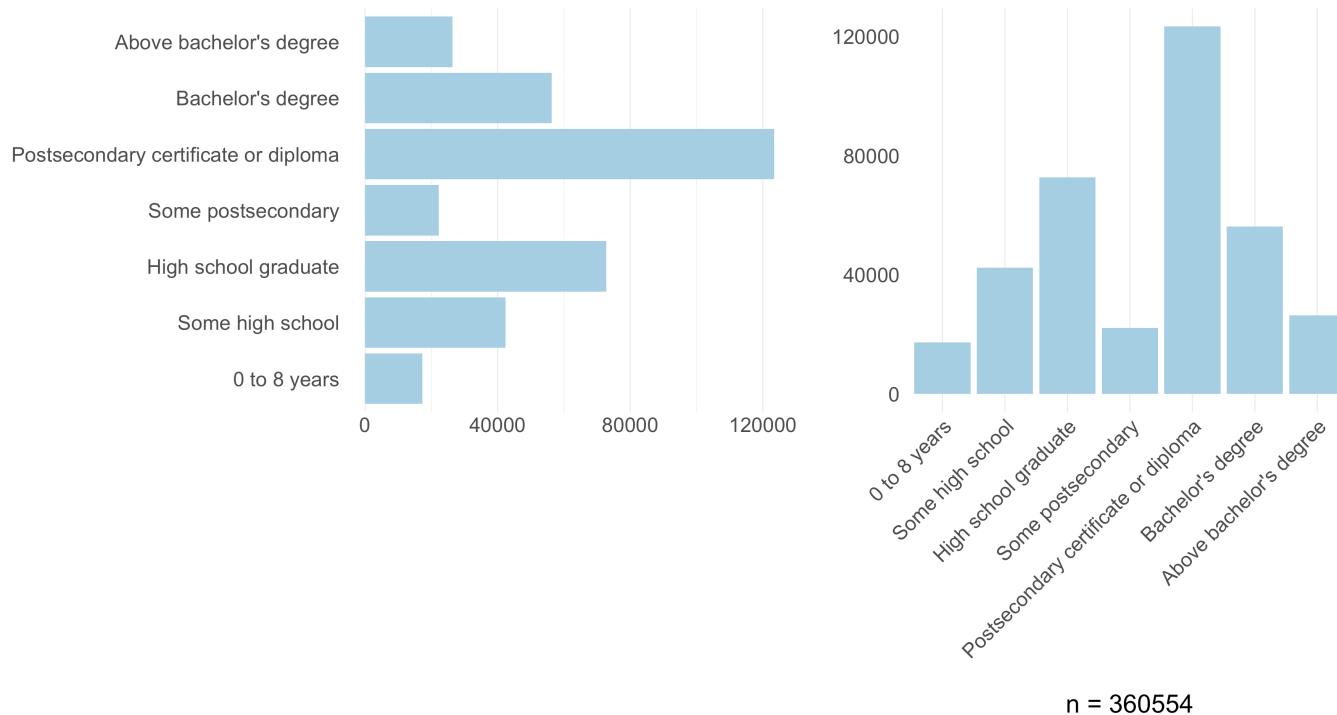
<b>Education</b>	<b>Count</b>
0 to 8 years	17299
Above bachelor's degree	26337
Bachelor's degree	56276
High school graduate	72725
Postsecondary certificate or diploma	123358
Some high school	42341
Some postsecondary	22218

# Labour Force Survey Data | Education

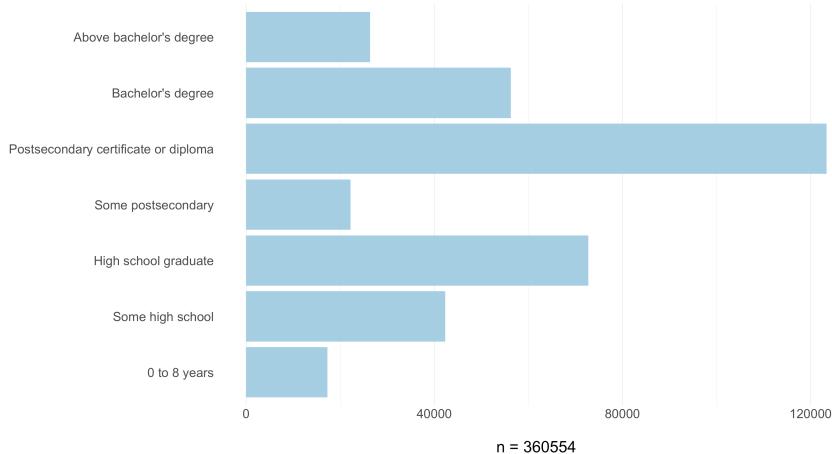


n = 360554

# Labeling Options

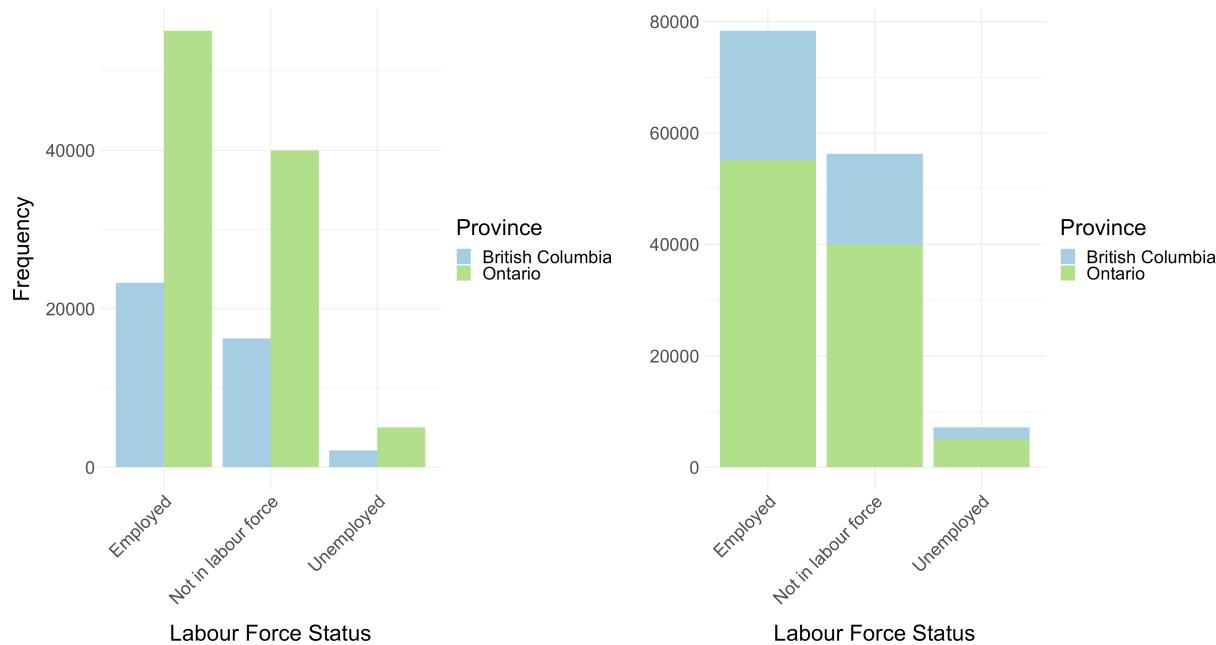


# Bar Chart Design Considerations



- 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**

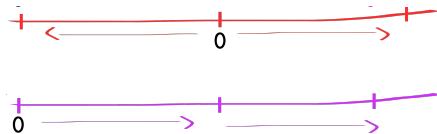
**Continuous** = **Measured**



# Numeric Data

**Interval = Greater or less than**

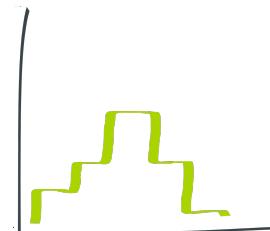
**Ratio = Percentage more or less**



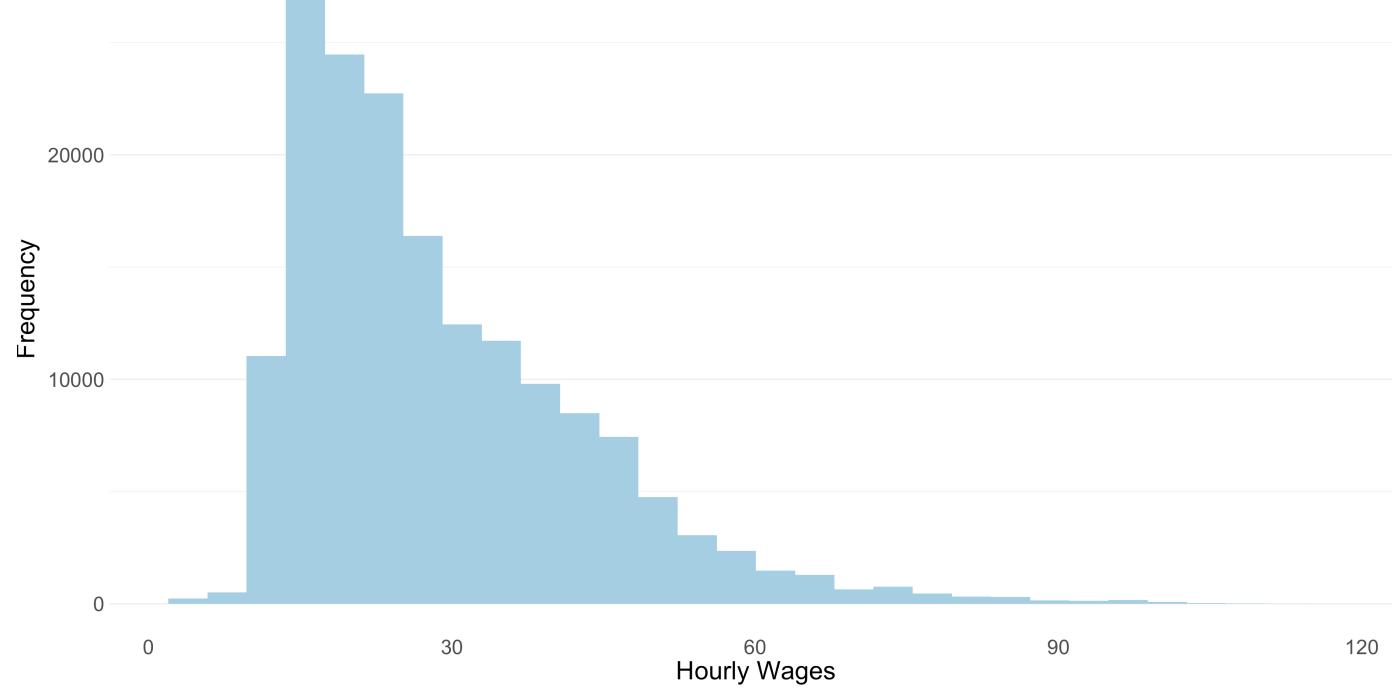
# **Counts of Numeric Data**

**Visualization tool**

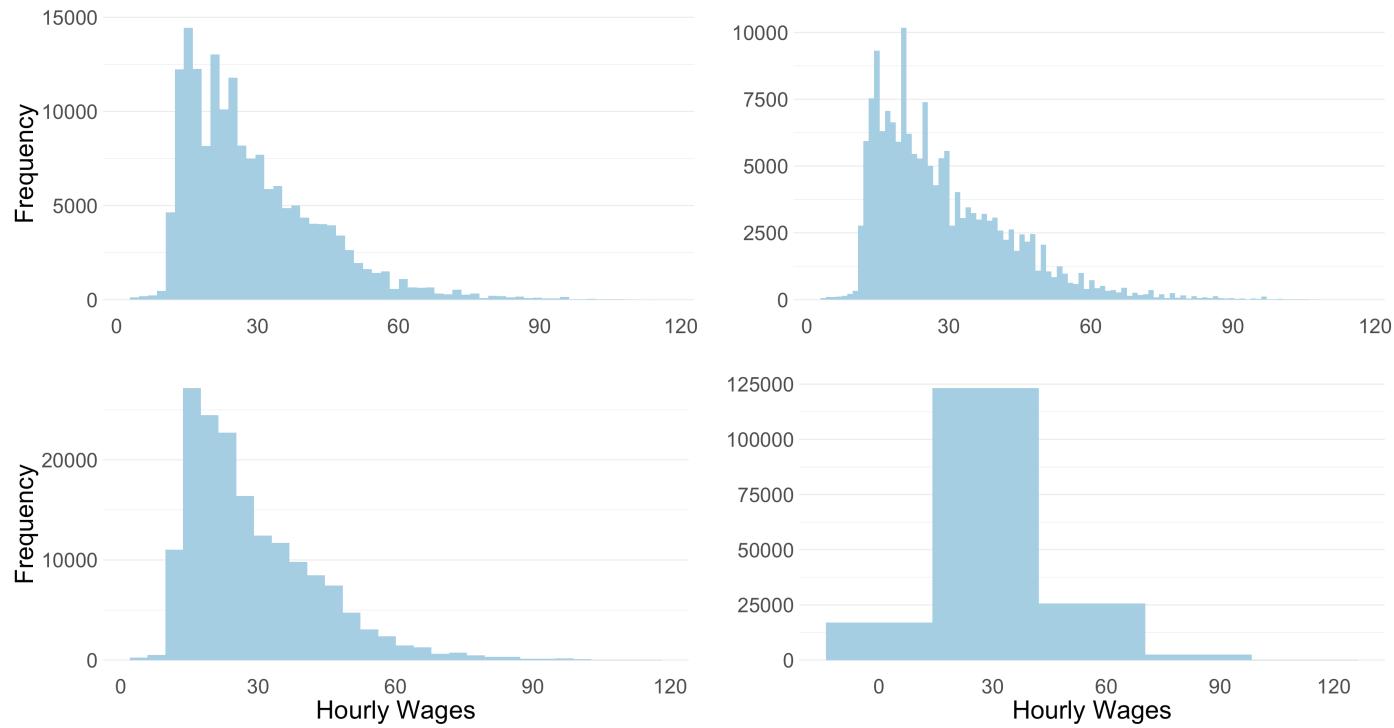
**Frequency plots with histograms**



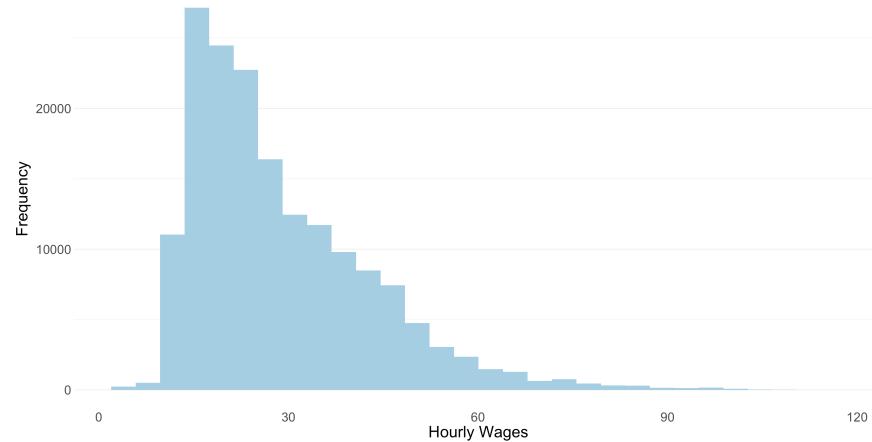
# Labour Force survey Data | Wages



# 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

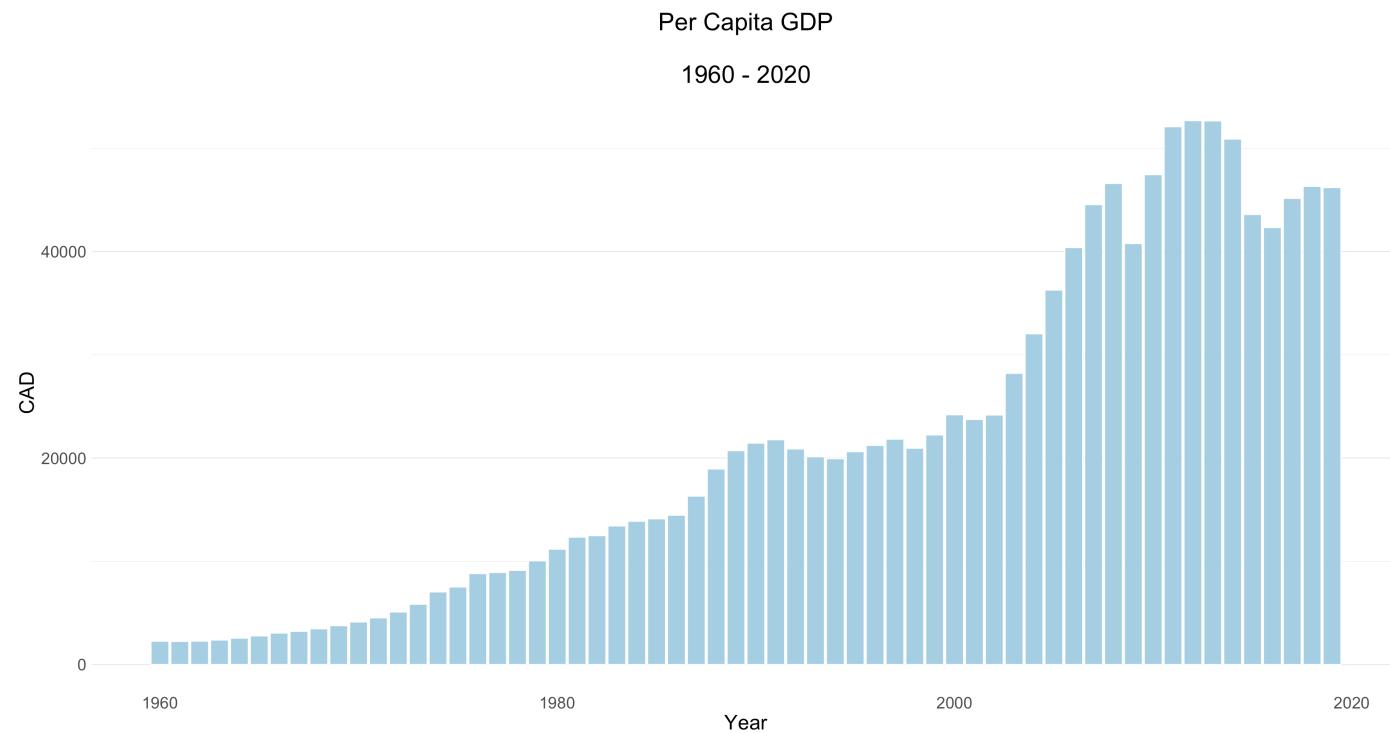
2011            52087.45

2010            47448.01

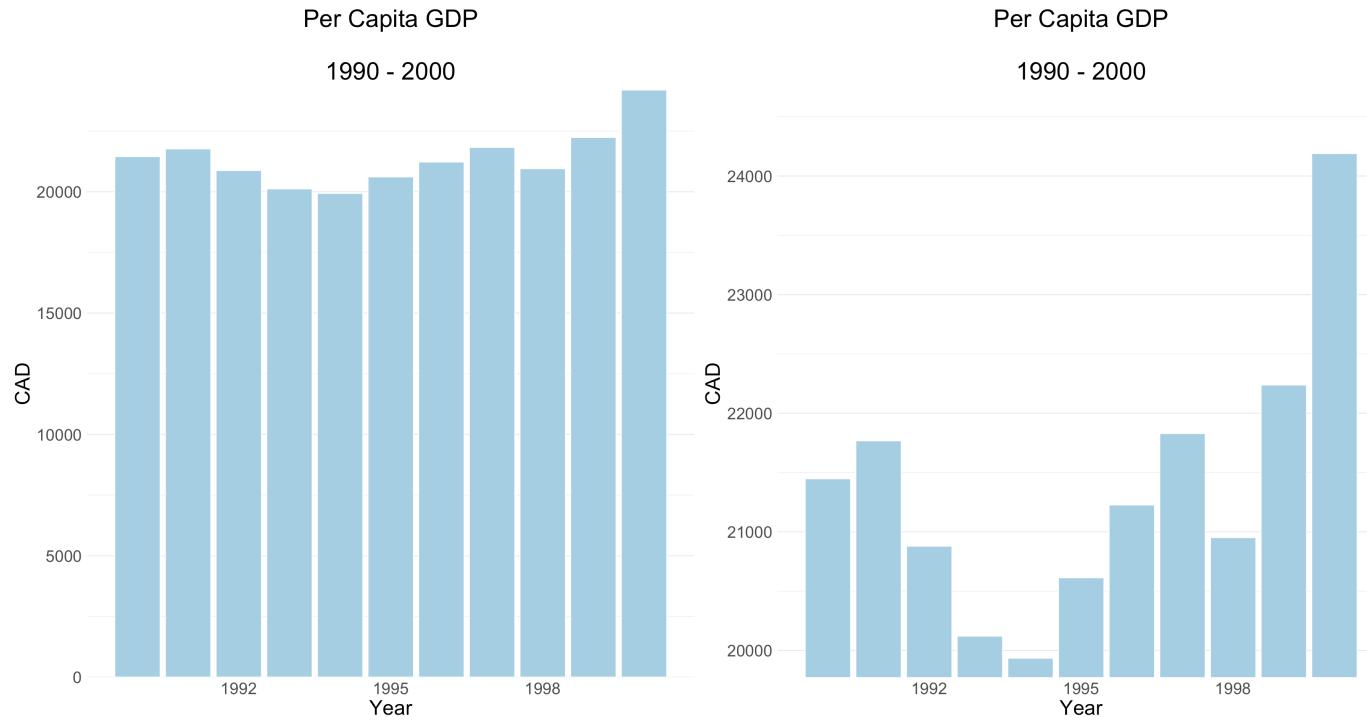
2009            40773.06

2008            46594.45

# Misleading Presentations



# Misleading Presentations

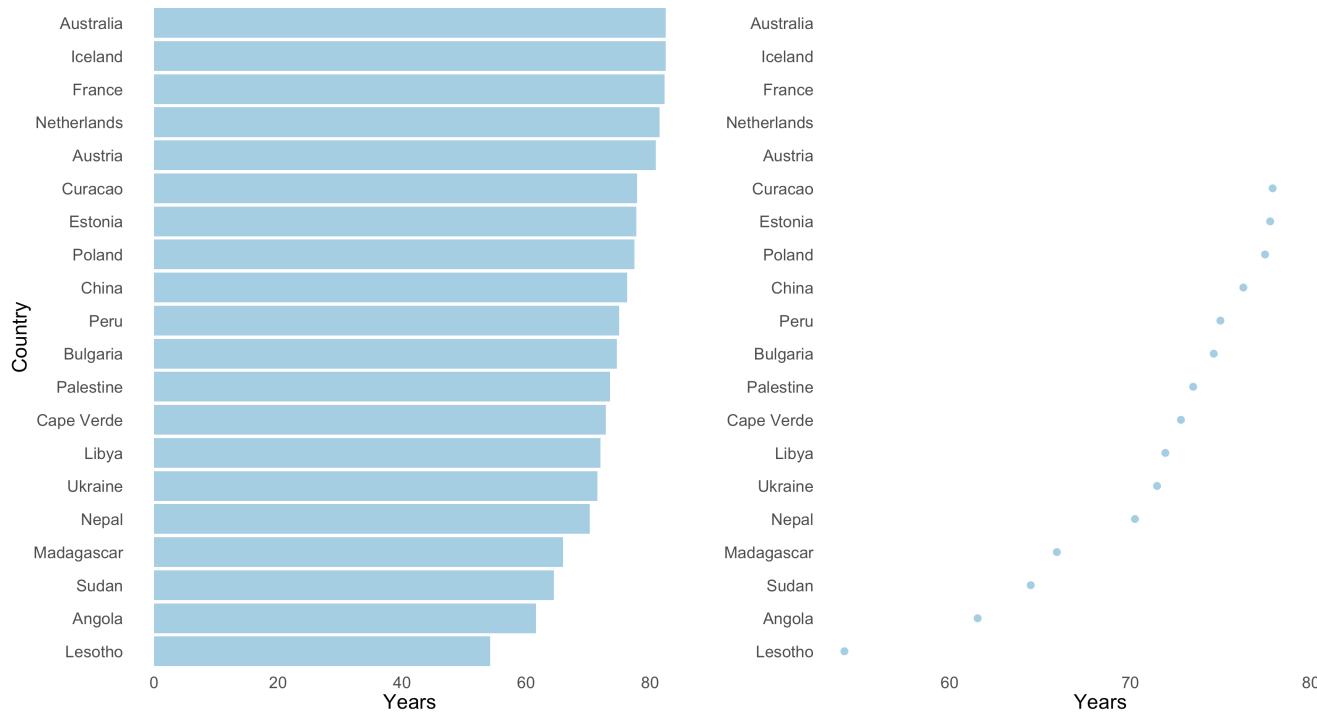


# **Dot Plots & Line Graphs**

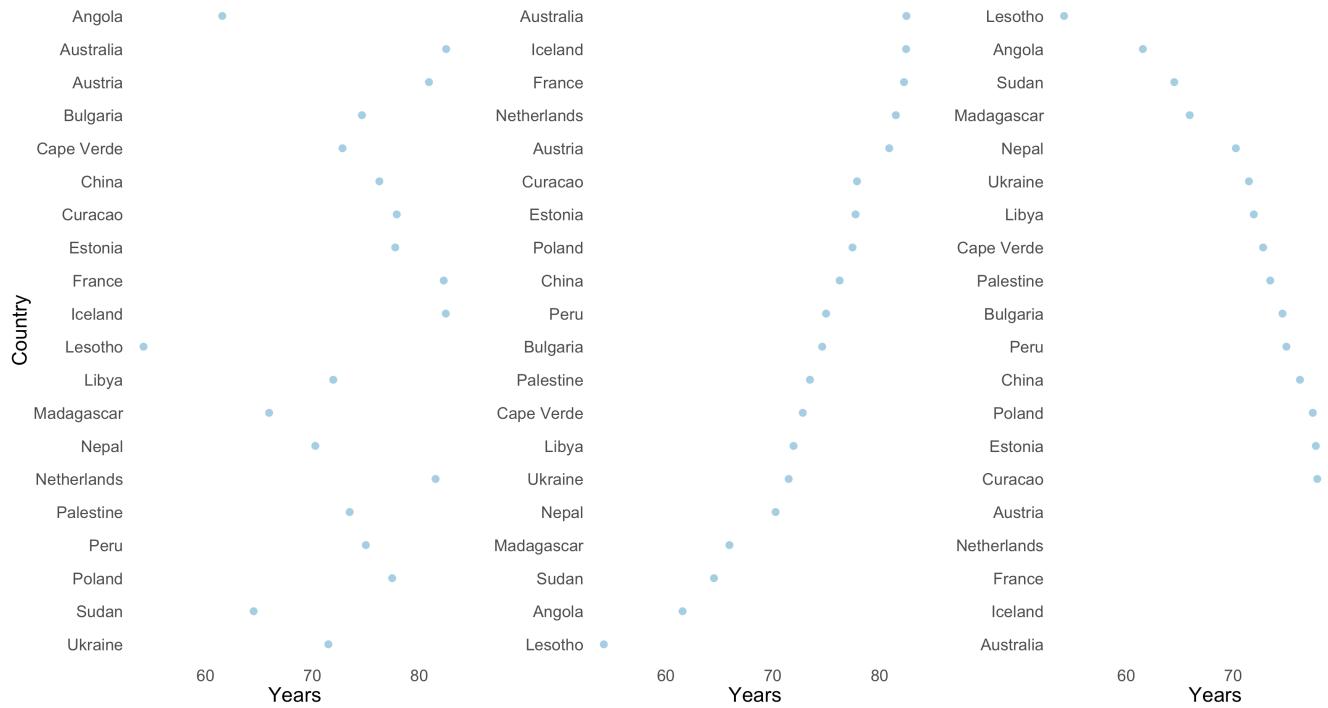
	<b>Country</b>	<b>Code</b>	<b>Year</b>	<b>Life.expectency</b>
7775	Peru	PER	2016	74.98300
2366	Curacao	CUW	2016	77.87317
1731	Cape Verde	CPV	2016	72.79800
5609	Lesotho	LSO	2016	54.17400
9941	Ukraine	UKR	2016	71.47634
9143	Sudan	SDN	2016	64.48600
456	Australia	AUS	2016	82.50000
228	Angola	AGO	2016	61.54700
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

	<b>Country</b>	<b>Code</b>	<b>Year</b>	<b>Life.expectency</b>
3373	France	FRA	2016	82.27317
5723	Libya	LBY	2016	71.93400
7547	Palestine	PSE	2016	73.47300
5974	Madagascar	MDG	2016	65.93200
1389	Bulgaria	BGR	2016	74.61463
6946	Netherlands	NLD	2016	81.50976

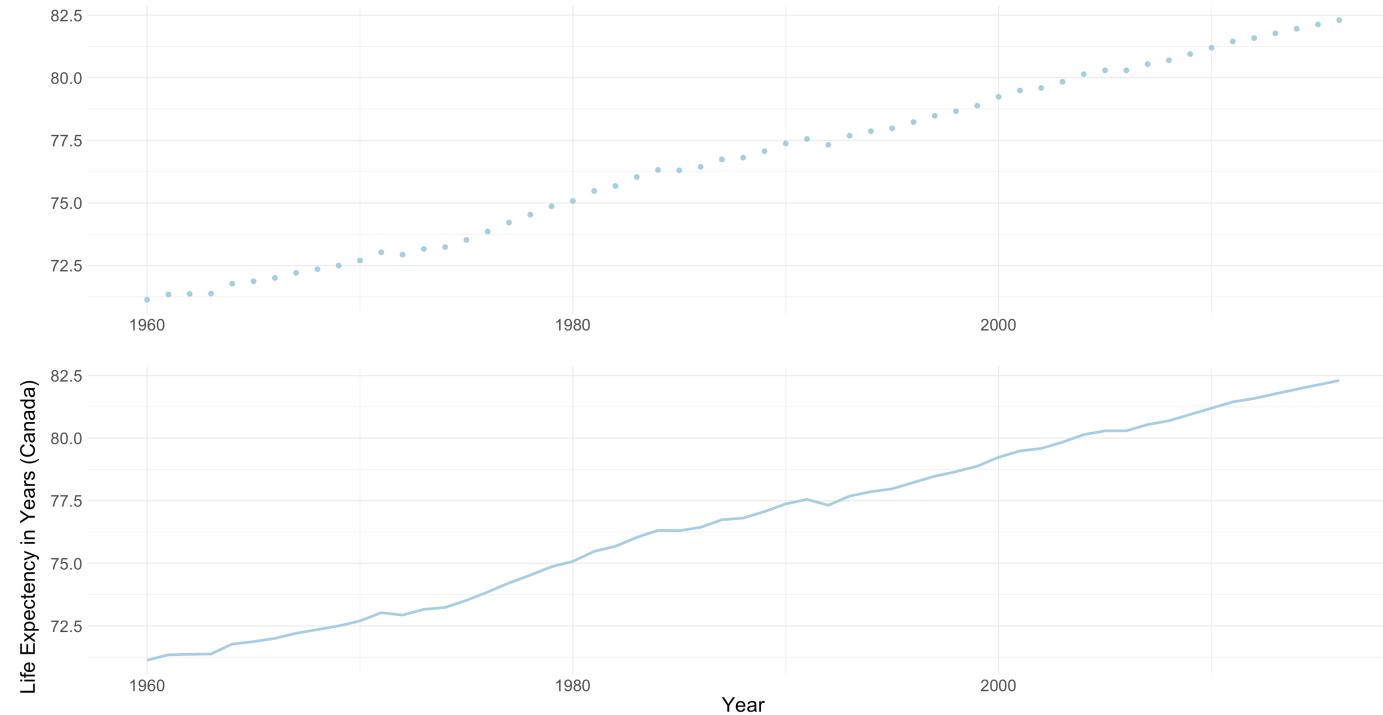
# Volume vs Points



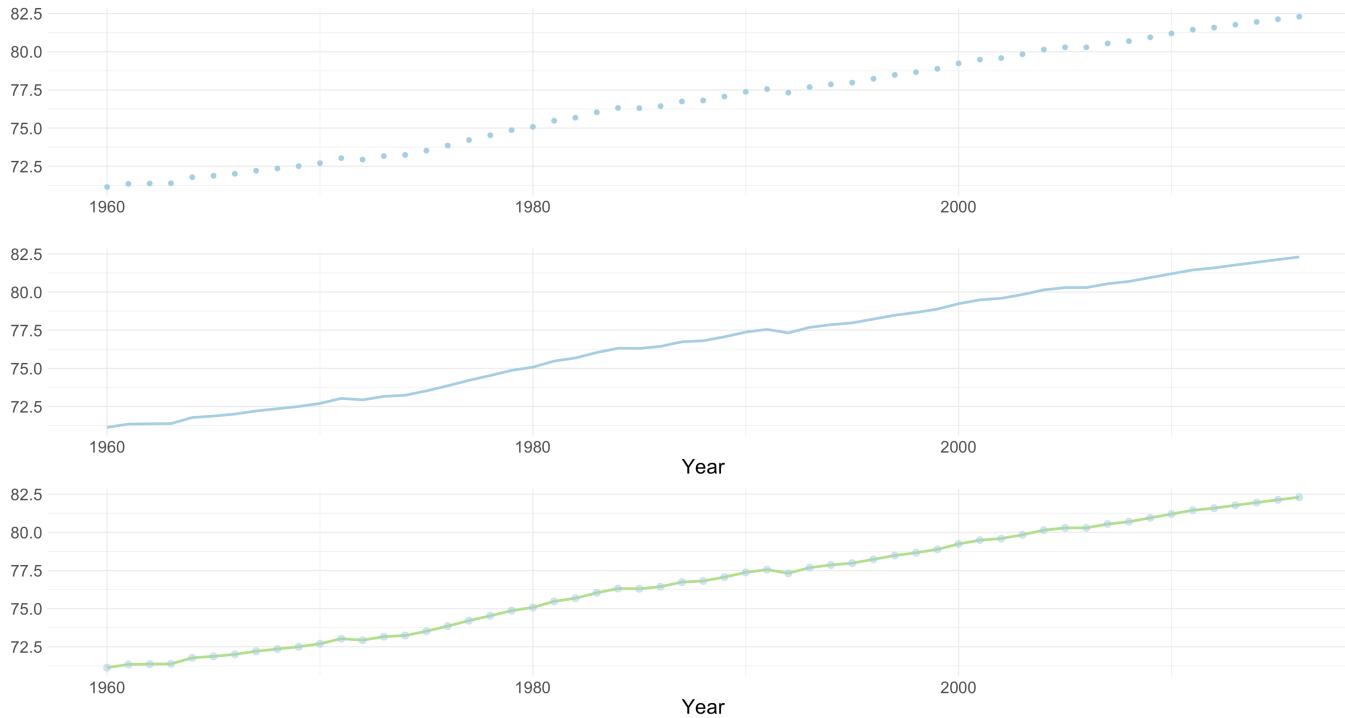
# Visual Order



# Time Sequences



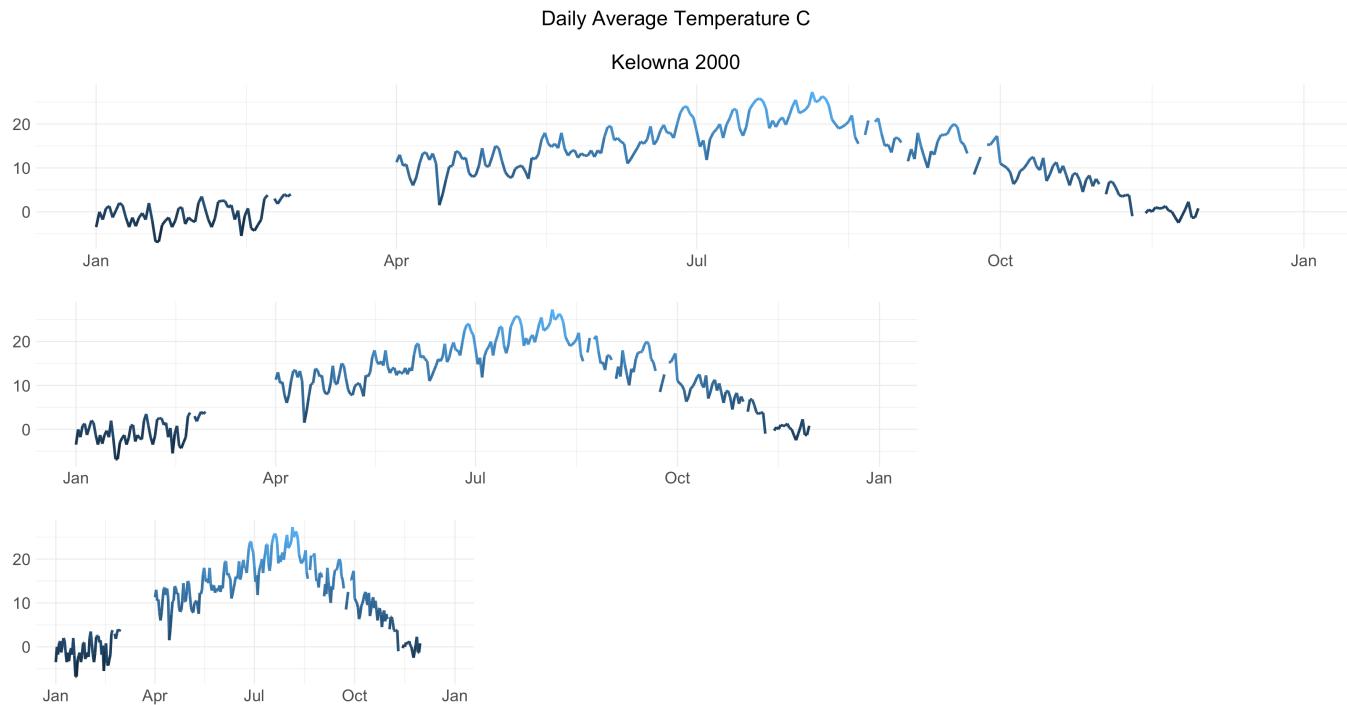
# Dots & Lines



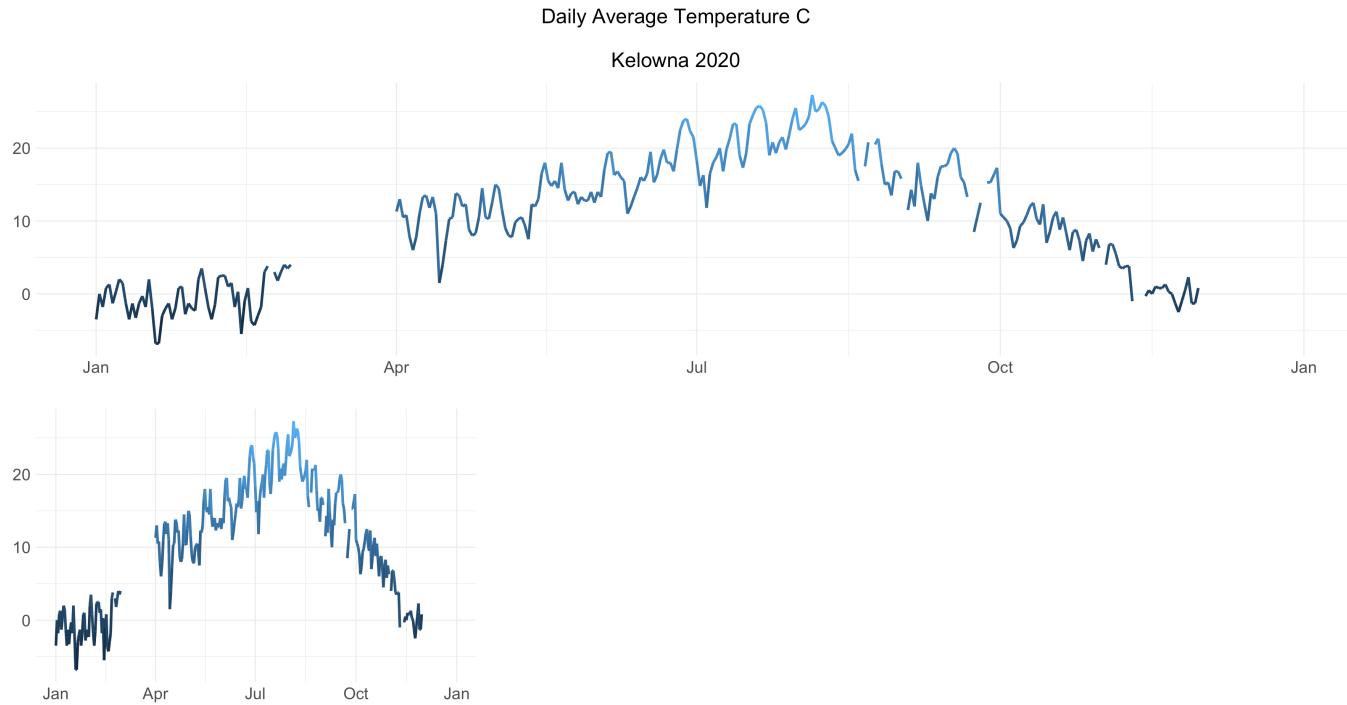
Longitude..x.	Latitude..y.	Station.Name	Climate.ID	Date.Time	Year	Month	Day	Data.Quality	Max.Temp..C.
-119.4	49.86	KELOWNA EAST	I123984	2000-01-01	2000			NA	-1.0
-119.4	49.86	KELOWNA EAST	I123984	2000-01-02	2000		2	NA	3.0
-119.4	49.86	KELOWNA EAST	I123984	2000-01-03	2000		3	NA	0.0
-119.4	49.86	KELOWNA EAST	I123984	2000-01-04	2000		4	NA	4.5
-119.4	49.86	KELOWNA EAST	I123984	2000-01-05	2000		5	NA	5.0
-119.4	49.86	KELOWNA EAST	I123984	2000-01-06	2000		6	NA	0.5
-119.4	49.86	KELOWNA EAST	I123984	2000-01-07	2000		7	NA	2.5
-119.4	49.86	KELOWNA EAST	I123984	2000-01-08	2000		8	NA	6.0
-119.4	49.86	KELOWNA EAST	I123984	2000-01-09	2000		9	NA	4.0

<b>Longitude..x.</b>	<b>Latitude..y.</b>	<b>Station.Name</b>	<b>Climate.ID</b>	<b>Date.Time</b>	<b>Year</b>	<b>Month</b>	<b>Day</b>	<b>Data.Quality</b>	<b>Max.Temp..C.</b>
-119.4	49.86	KELOWNA EAST	1123984	2000-01-10	2000	1	10	NA	2.5
-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

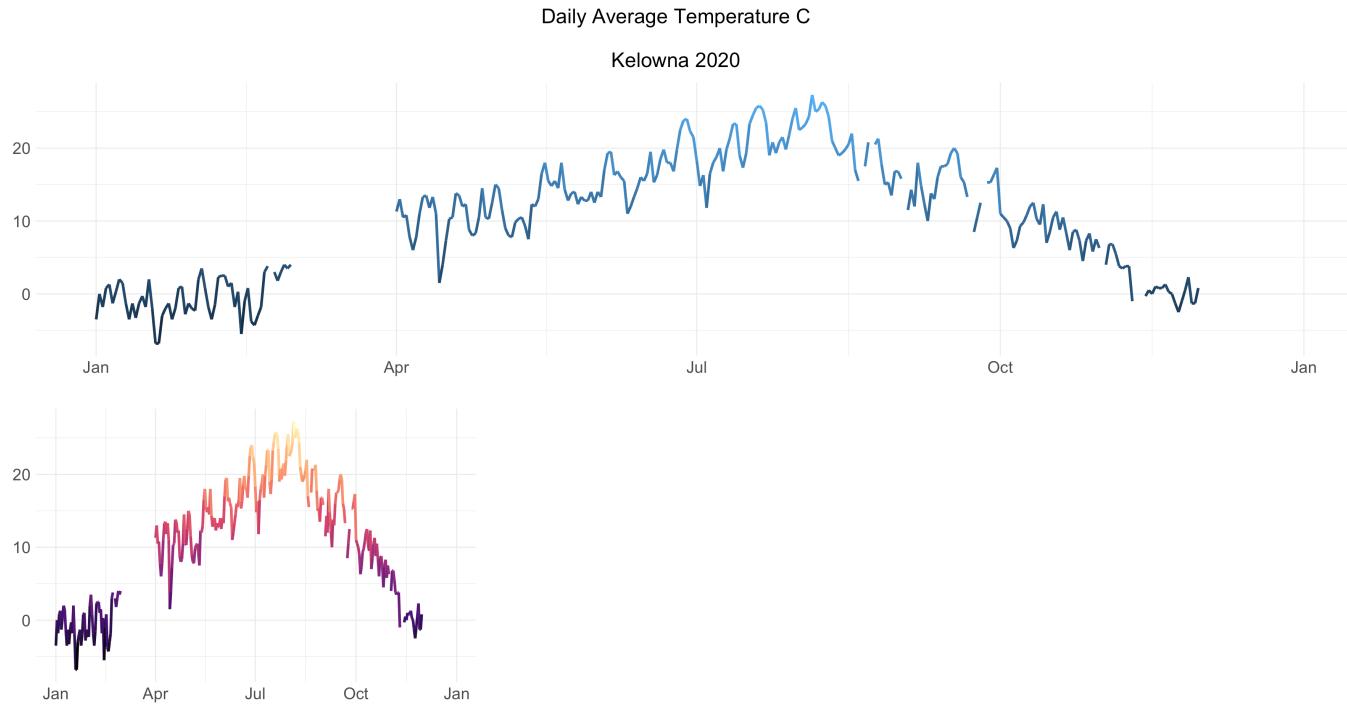
# Misleading Presentations



# Misleading Presentations



# Misleading Presentations



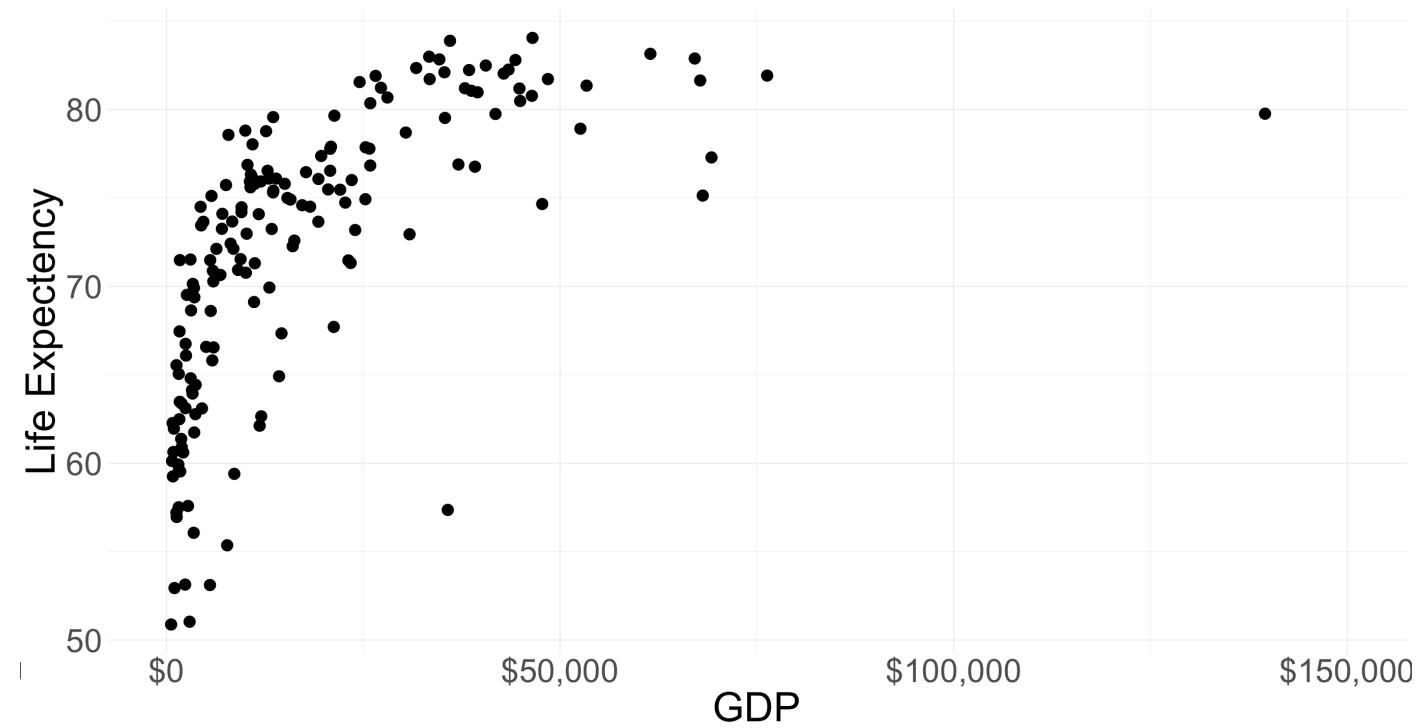
# **Layering Data and Statistics**

<b>Country</b>	<b>Code</b>	<b>Year</b>	<b>Population</b>	<b>Continent</b>	<b>Life.Expectency</b>	<b>GDP</b>
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

<b>Country</b>	<b>Code</b>	<b>Year</b>	<b>Population</b>	<b>Continent</b>	<b>Life.Expectency</b>	<b>GDP</b>
Vietnam	VNM	2015	92677000	Asia	75.110	5733
Congo	COG	2015	4856000	Africa	63.097	4526
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

<b>Country</b>	<b>Code</b>	<b>Year</b>	<b>Population</b>	<b>Continent</b>	<b>Life.Expectency</b>	<b>GDP</b>
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

<b>Country</b>	<b>Code</b>	<b>Year</b>	<b>Population</b>	<b>Continent</b>	<b>Life.Expectency</b>	<b>GDP</b>
Vietnam	VNM	2015	92677000	Asia	75.110	5733
Congo	COG	2015	4856000	Africa	63.097	4526
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

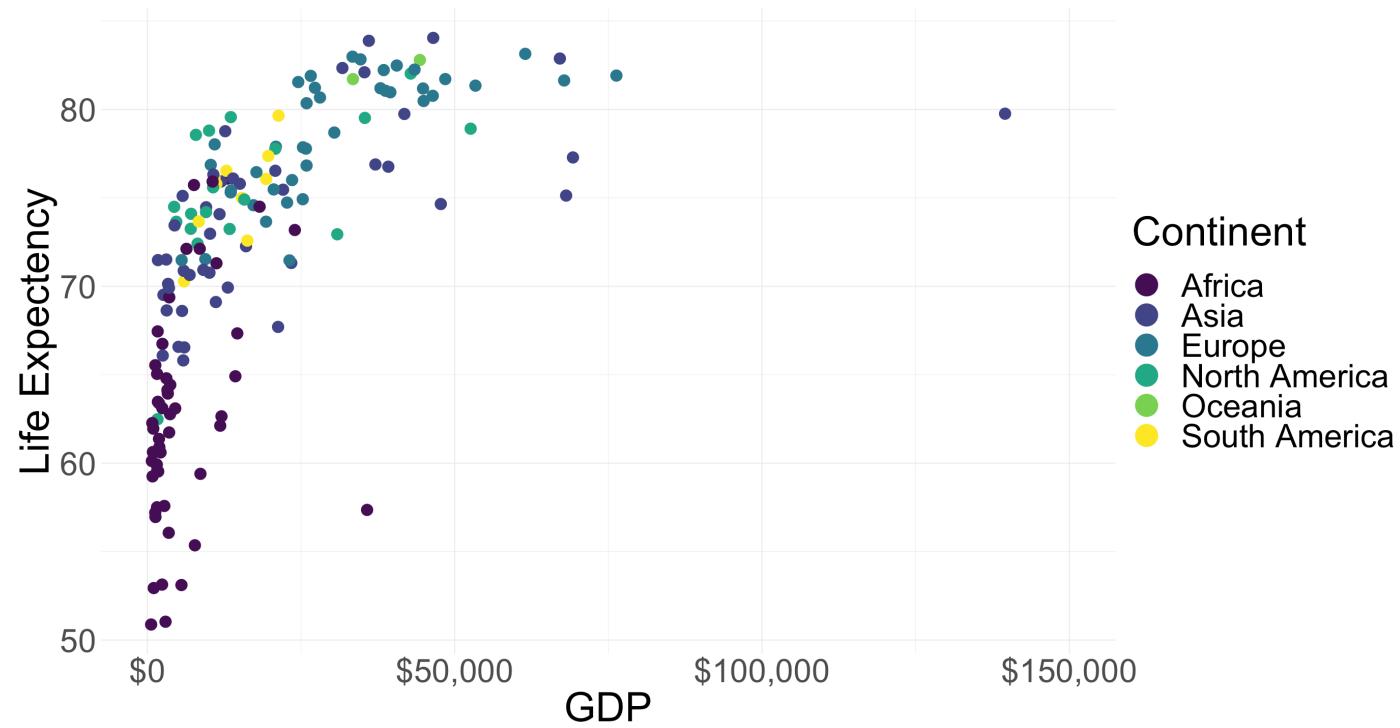


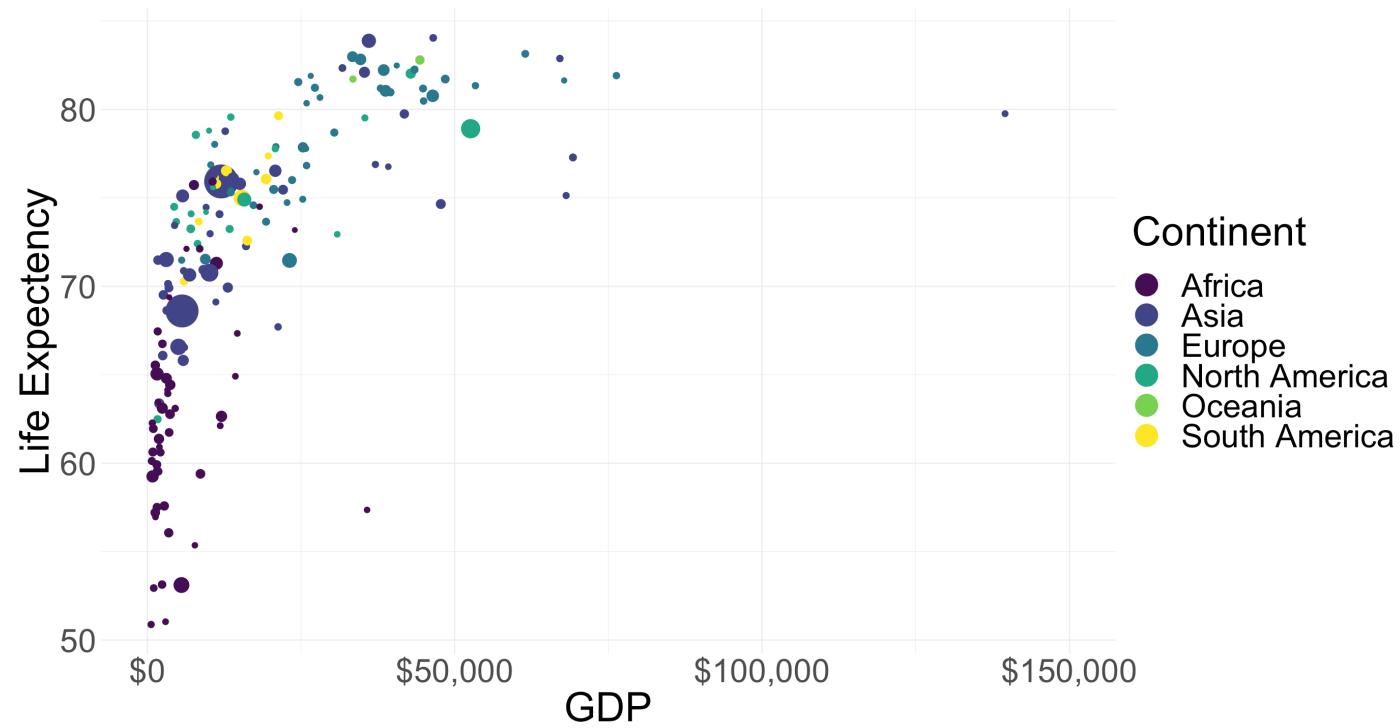
<b>Country</b>	<b>Code</b>	<b>Year</b>	<b>Population</b>	<b>Continent</b>	<b>Life.Expectency</b>	<b>GDP</b>
Lesotho	LSO	2015	2059000	Africa	51.038	2954
Armenia	ARM	2015	2926000	Asia	74.467	9552
Uruguay	URY	2015	3412000	South America	77.369	19668
Slovakia	SVK	2015	5436000	Europe	76.827	25896
Bosnia and 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

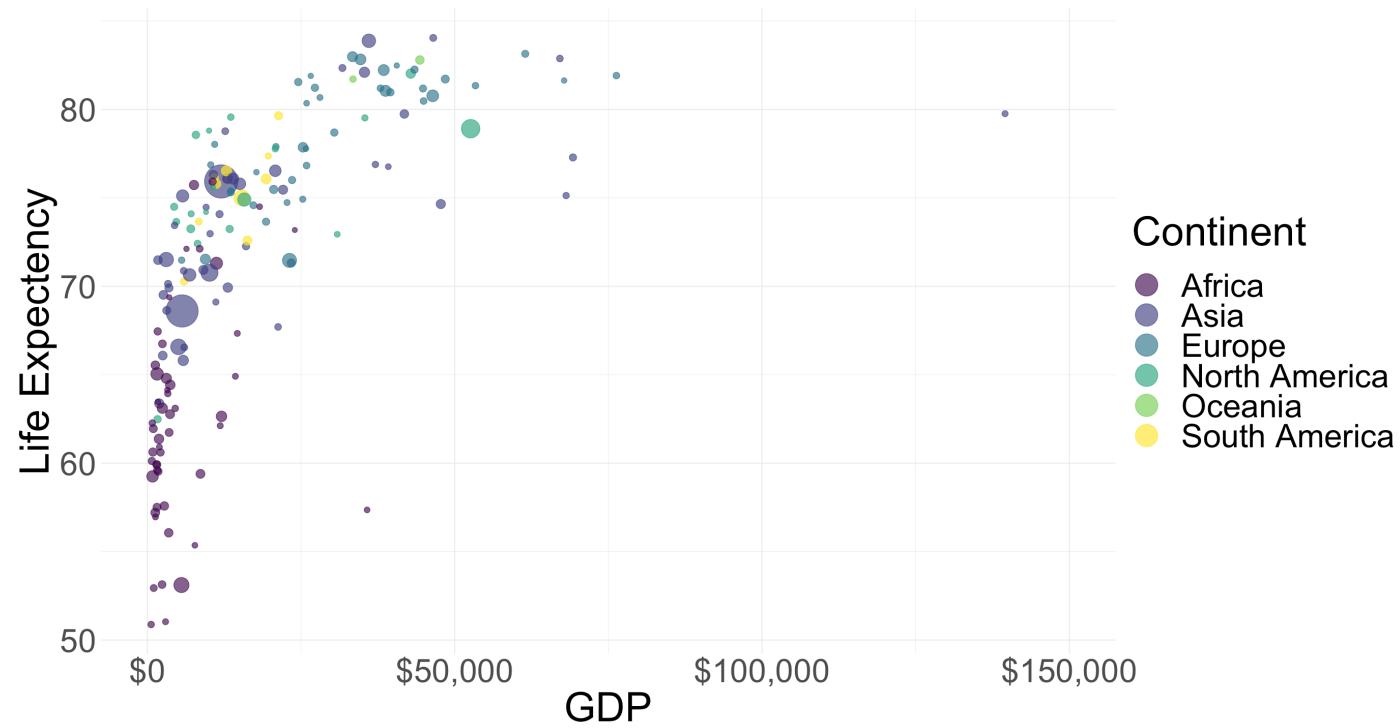
<b>Country</b>	<b>Code</b>	<b>Year</b>	<b>Population</b>	<b>Continent</b>	<b>Life.Expectency</b>	<b>GDP</b>
Vietnam	VNM	2015	92677000	Asia	75.110	5733
Congo	COG	2015	4856000	Africa	63.097	4526
Gabon	GAB	2015	1948000	Africa	64.913	14315
Nepal	NPL	2015	27015000	Asia	69.515	2607
North Macedonia	MKD	2015	2079000	Europe	75.406	13586
Cyprus	CYP	2015	1161000	Europe	80.350	25903

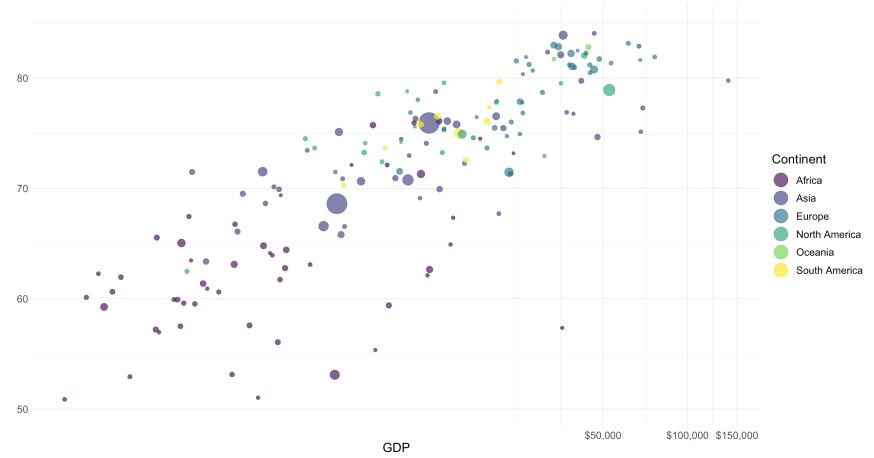
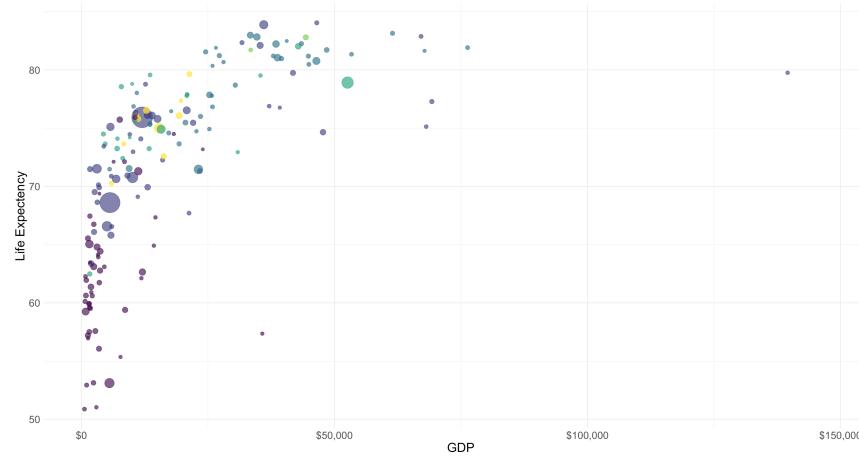
<b>Country</b>	<b>Code</b>	<b>Year</b>	<b>Population</b>	<b>Continent</b>	<b>Life.Expectency</b>	<b>GDP</b>
Lesotho	LSO	2015	2059000	Africa	51.038	2954
Armenia	ARM	2015	2926000	Asia	74.467	9552
Uruguay	URY	2015	3412000	South America	77.369	19668
Slovakia	SVK	2015	5436000	Europe	76.827	25896
Bosnia and 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

<b>Country</b>	<b>Code</b>	<b>Year</b>	<b>Population</b>	<b>Continent</b>	<b>Life.Expectency</b>	<b>GDP</b>
Vietnam	VNM	2015	92677000	Asia	75.110	5733
Congo	COG	2015	4856000	Africa	63.097	4526
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



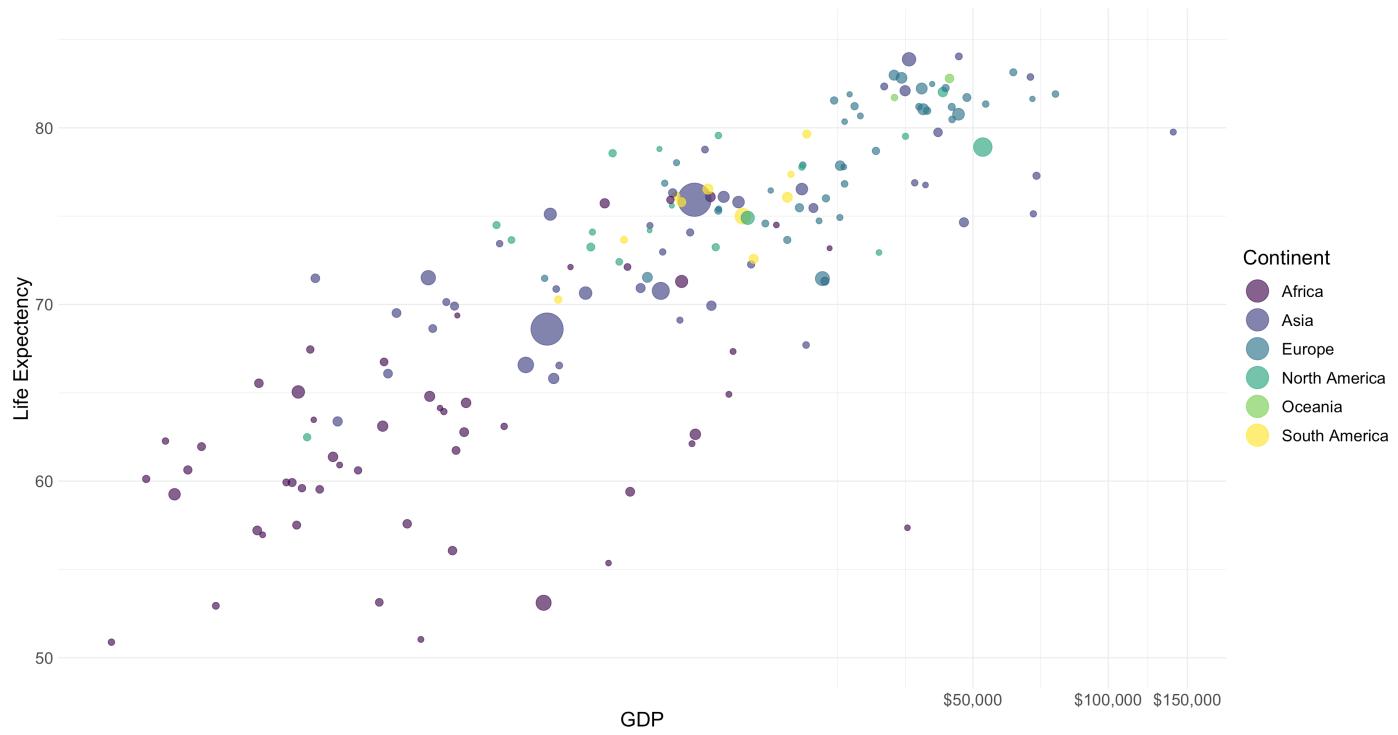




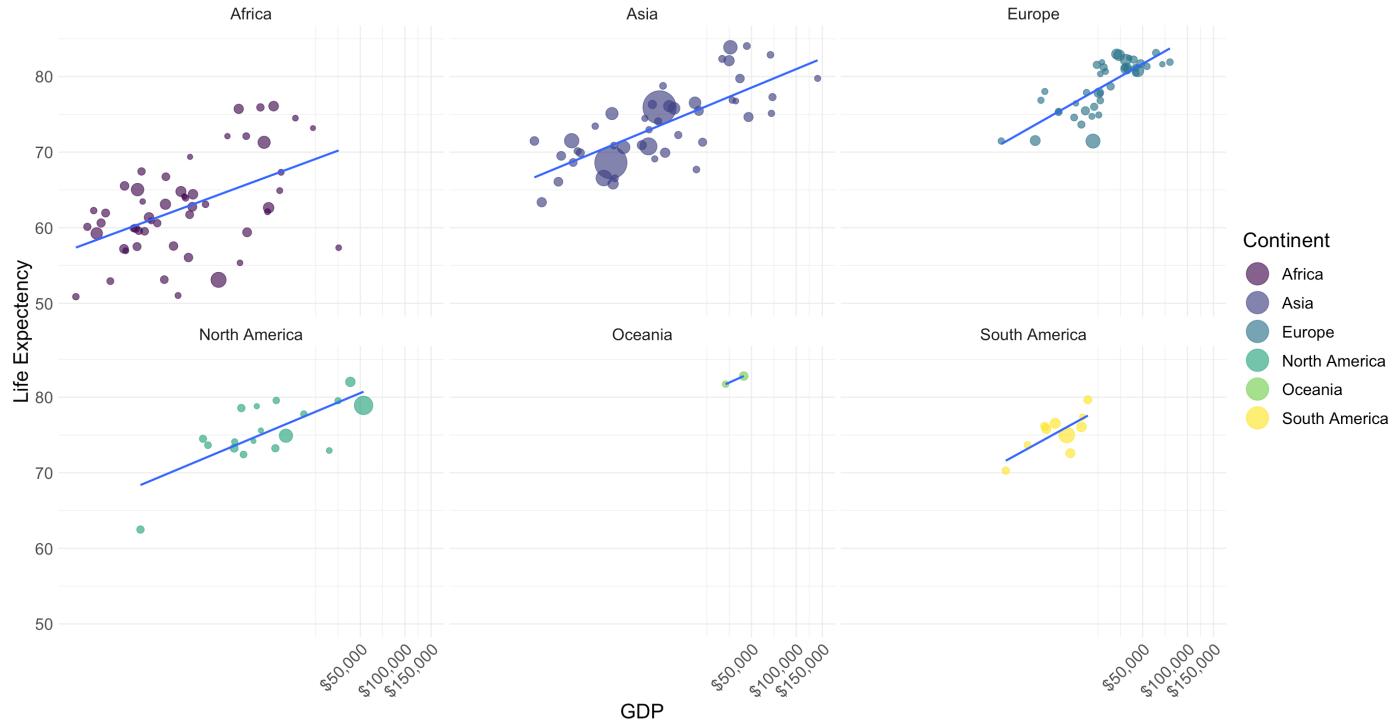


Continent

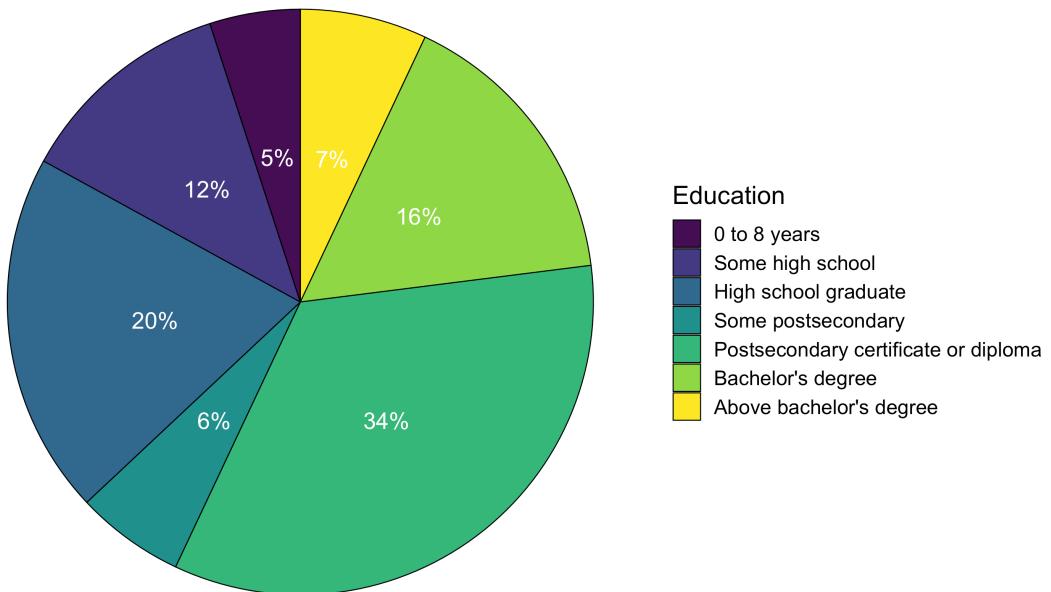
- Africa
- Asia
- Europe
- North America
- Oceania
- South America

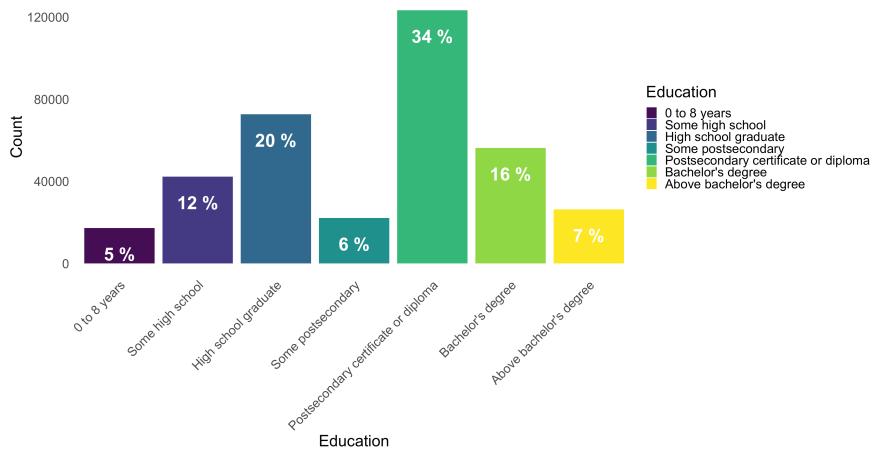
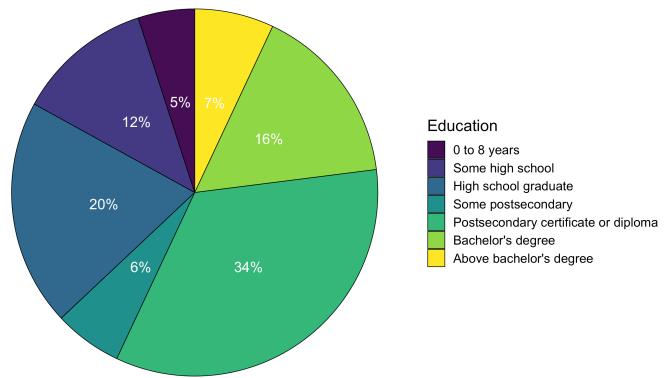






# Pie Charts





# **Finding Data**

# CBC Fatal Police Encounters



**2020 already a particularly deadly year for people killed in police encounters, CBC research shows**

CBC's Deadly Force database looks at role of race, mental health in deaths

Inayat Singh  
CBC News • July 23, 2020

# **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>

## **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