An Introduction to Data Visualization With Tableau

November 18, 2020

Devon Mordell, Educational Developer
MacPherson Institute, McMaster University

Do More With Digital Scholarship Series

Lewis & Ruth Sherman Centre for Digital Scholarship
scds.ca

McMaster University Library

Image: Mhsheikholeslami (CC 4.0 BY-SA)

Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences.

In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

Code of Conduct

The Sherman Centre and the McMaster University Library are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

Please refer to our code of conduct webpage for more information:

scds.ca/events/code-of-conduct/

We’re here to help!

Use **TH:** [query] in chat to let the facilitators know that you’re having technical issues

By the end of this workshop…

You'll be able to:

- Identify what kinds of tasks Tableau is suited to
- Explain the concept of data provenance
- Use visualizations in Tableau to make observations about a dataset
- Create a customized dashboard in Tableau

Have you downloaded Tableau?

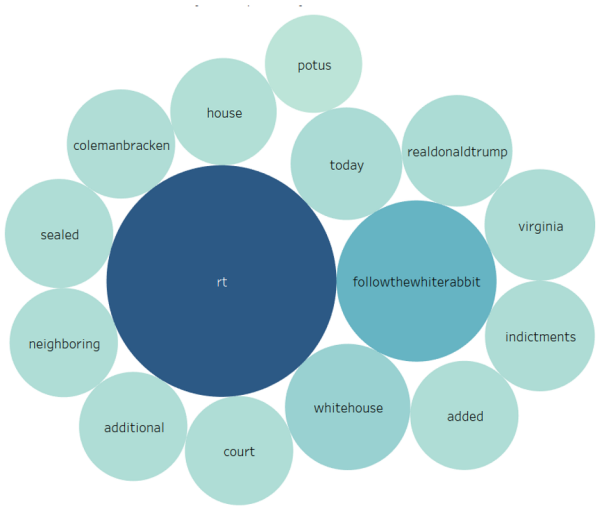
- If not, available options:
  - Tableau Public (limited & public)
  - 14-day trial (use for now)
  - 1-year license for students & instructors*

Tableau is designed for…

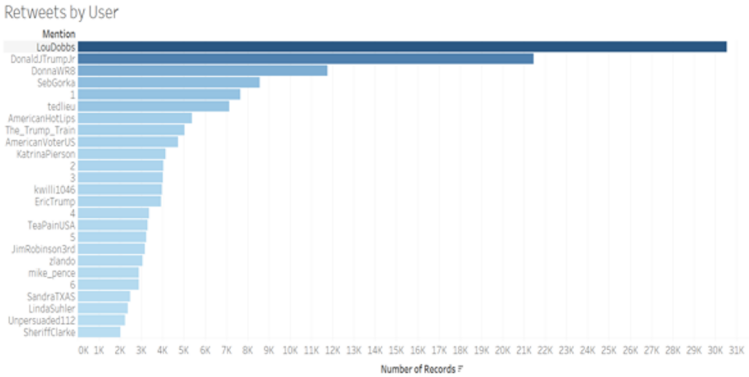
Visualizing **tabular data**

Data item	Attribute 1 - Location	Attribute 2 - Tonnes CO_2e
[Measurement at Facility 1]	Alberta	2,000,000
[Measurement at Facility 2]	Ontario	600,000

Textual analysis?



Social network analysis?



A brief primer on data visualization…

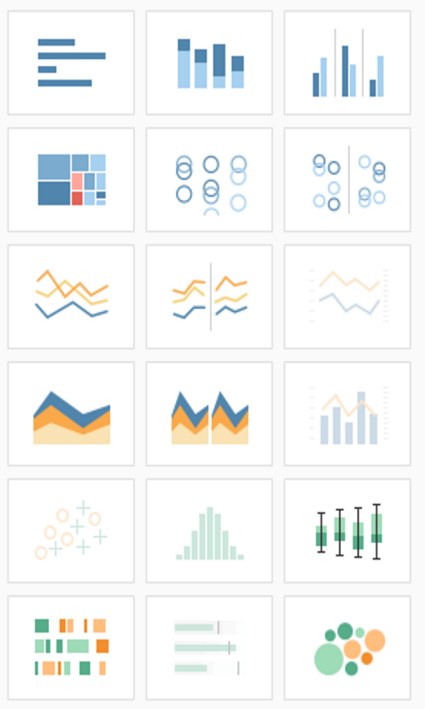
Data Visualization Primer

Idioms

“A [visualization] idiom is a distinct approach to creating and manipulating visual representations”

(Munzner 2014)

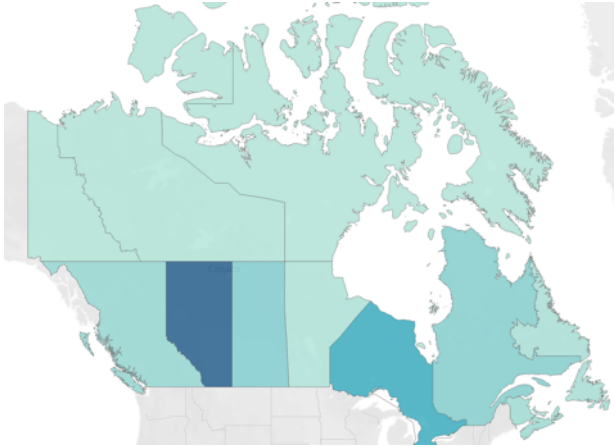
I.e. “type” of visualization used to represent data



Data Visualization Primer

Use **channels** to represent different **attributes** of the data...

- Colour
- Size
- Area
- Shape



Data Visualization Primer

Categorical (often qualitative):

e.g. Facility name, Industry, Province, Year

Continuous (always quantitative):

e.g. Latitude, CO_2 emissions (in tonnes)

Greenhouse Gas (GHG) Emissions in Canada, 2004-2017

The dataset

On Data Provenance

I.e. record where the data came from and what kinds of transformations were performed on it: view original dataset

As with any other research,

documentation → trustworthiness

Exploratory data analysis (EDA)

What do the columns mean?

- CH_4 (tonnes CO_2e) = CH_4 (tonnes) * 25
- N_2O (tonnes CO_2e) = N_2O (tonnes) * 298

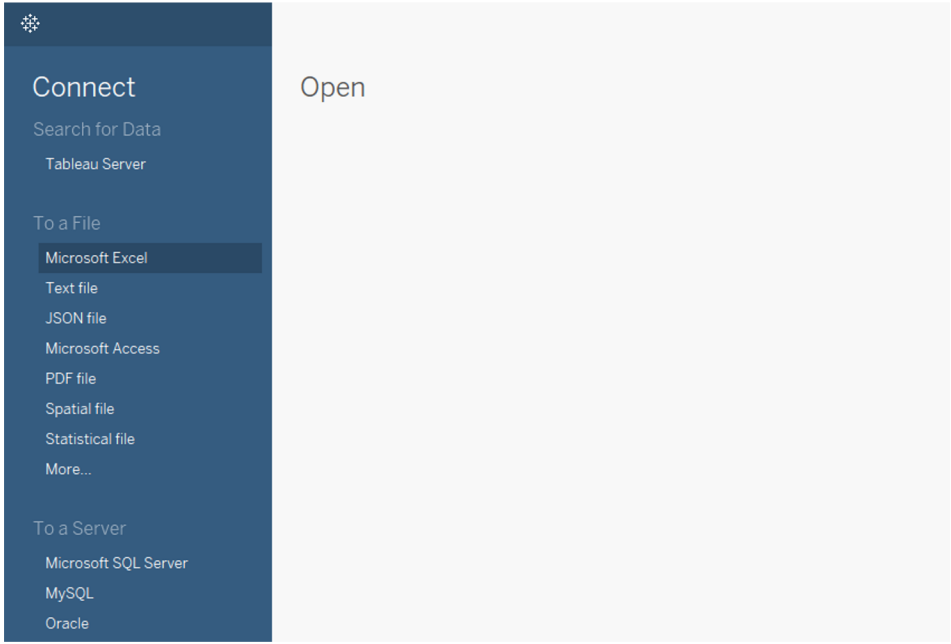
Any data collection or transcription errors?

EDA is an iterative process…

↑ Surface data collection errors & ‘messy’ data

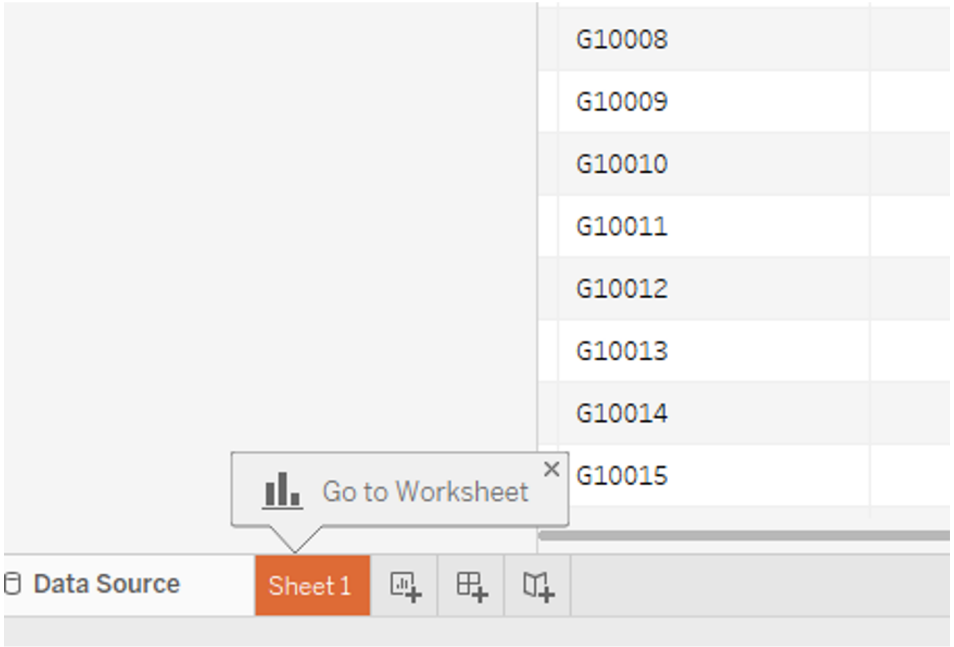
↑ Test assumptions, detect variables & refine question

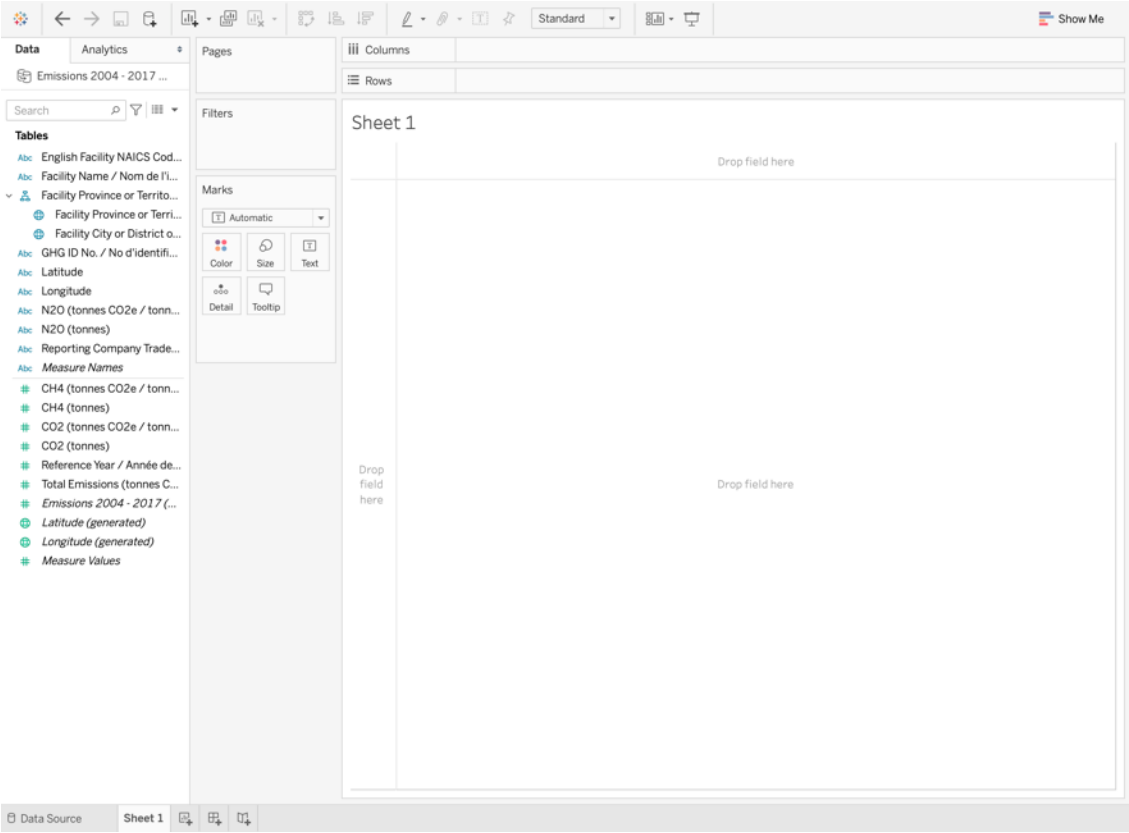
1. Connect to a data source (Excel, CSV, JSON, PDF, etc.)



2. Select & drag the sheet containing the data to be visualized

3. Go to worksheet...



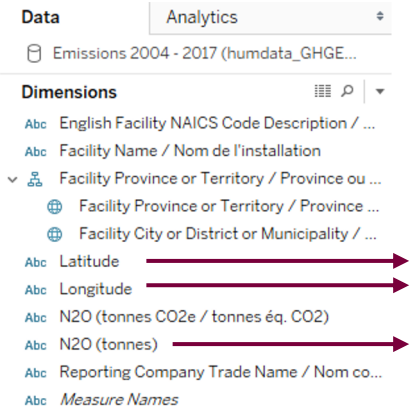

Standard
Show Me
Data
Analytics
Emissions 2004 - 2017 ...
Search
Tables
English Facility NAICS Cod...
Facility Name / Nom de l'i...
Facility Province or Territo...
Facility Province or Terri...
Facility City or District o...
GHG ID No. / No d'identifi...
Latitude
Longitude
N2O (tonnes CO2e / tonn...
N2O (tonnes)
Reporting Company Trade...
Measure Names
CH4 (tonnes CO2e / tonn...
CH4 (tonnes)
CO2 (tonnes CO2e / tonn...
CO2 (tonnes)
Reference Year / Année de...
Total Emissions (tonnes C...
Emissions 2004 - 2017 (...
Latitude (generated)
Longitude (generated)
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Text
Detail
Tooltip
Columns
Rows
Sheet 1
Drop field here
Drop field here
Drop field here
Data Source
Sheet 1

Dimensions vs Measures

Dimensions: qualitative values that can be used to categorize and segment data
→ **~categorical data attributes**

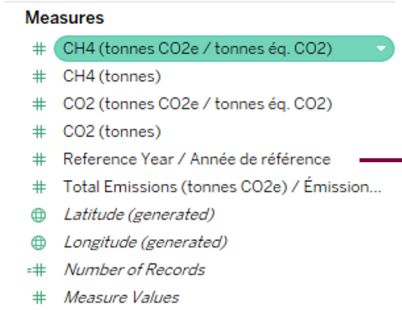
Measures: quantitative values that can be aggregated (e.g. sum)
→ **~continuous data attributes**

4. Verify and correct your data types

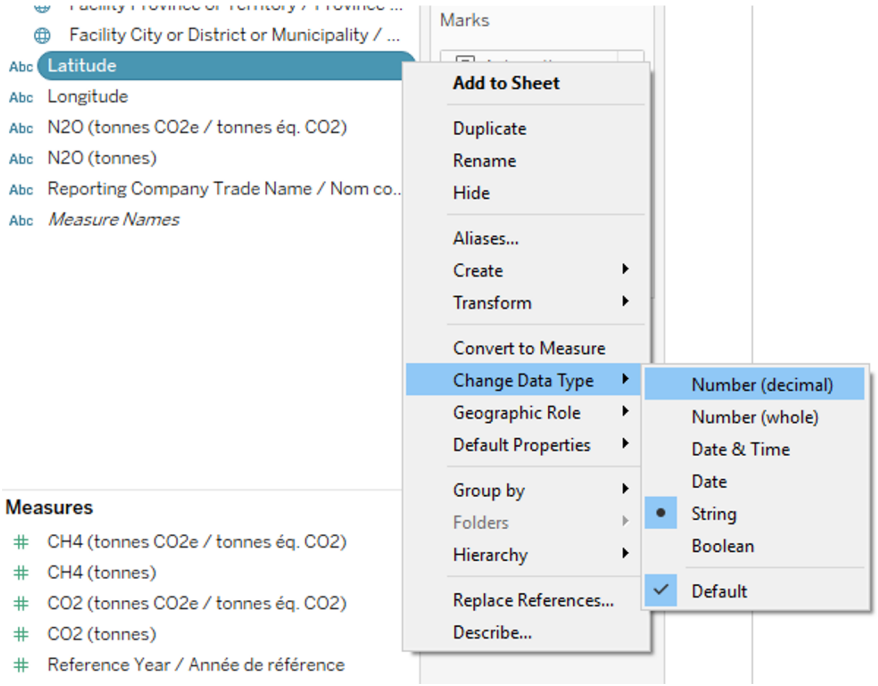


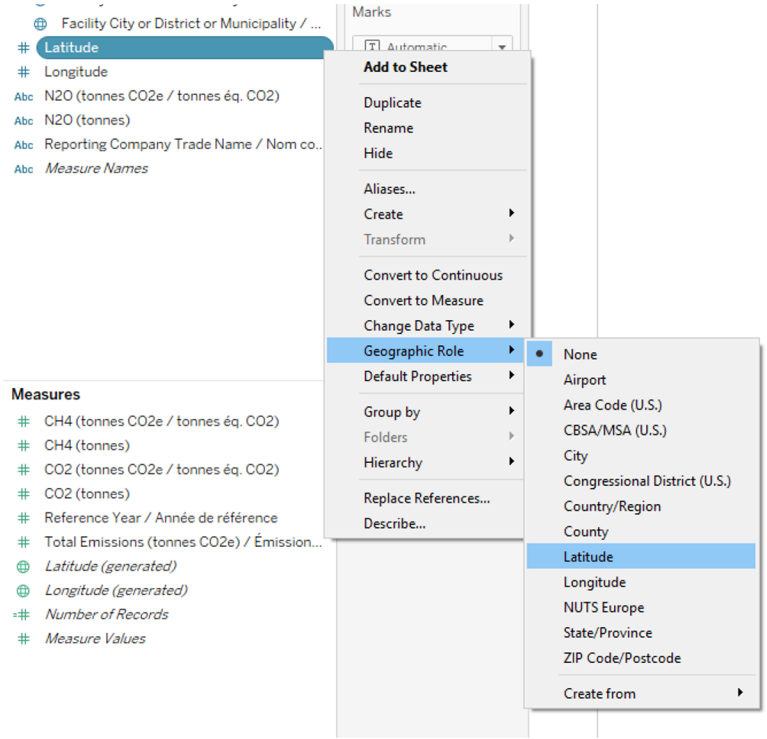
geographic role

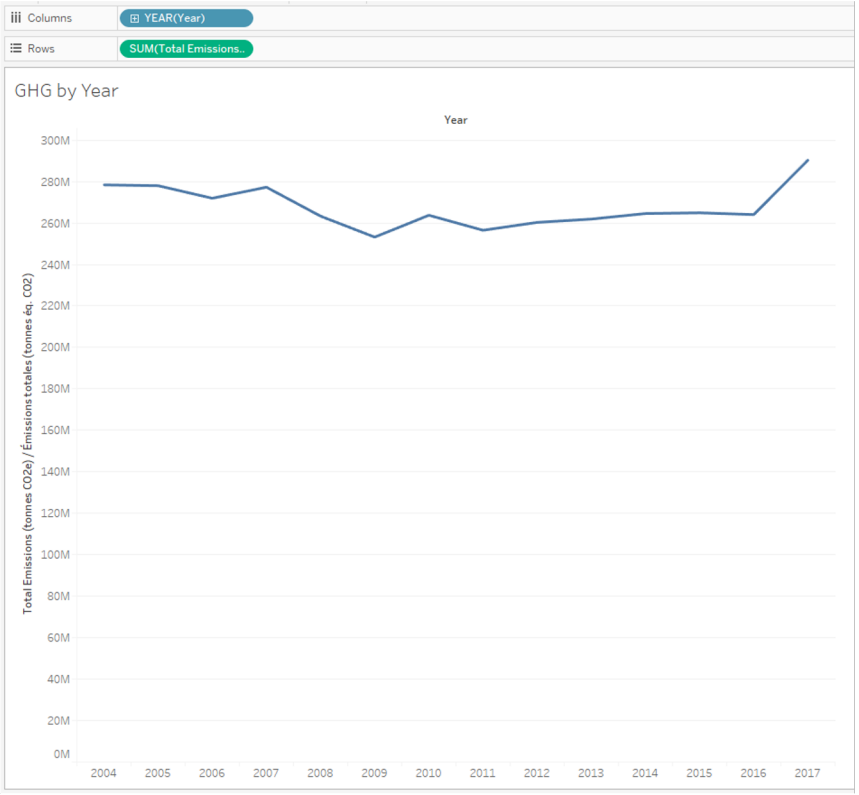
numerical

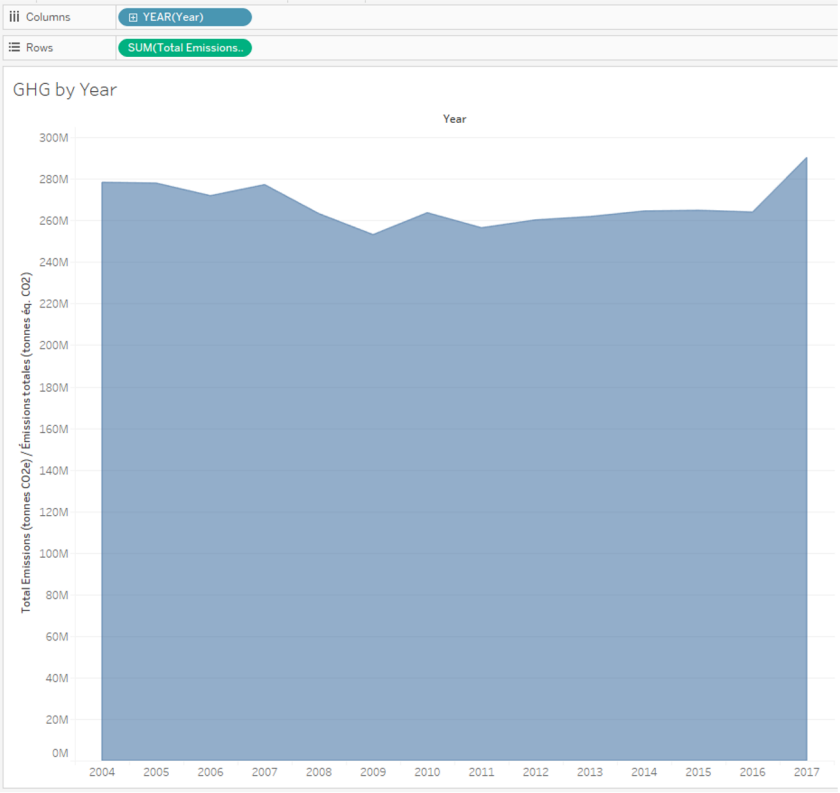


categorical, date role

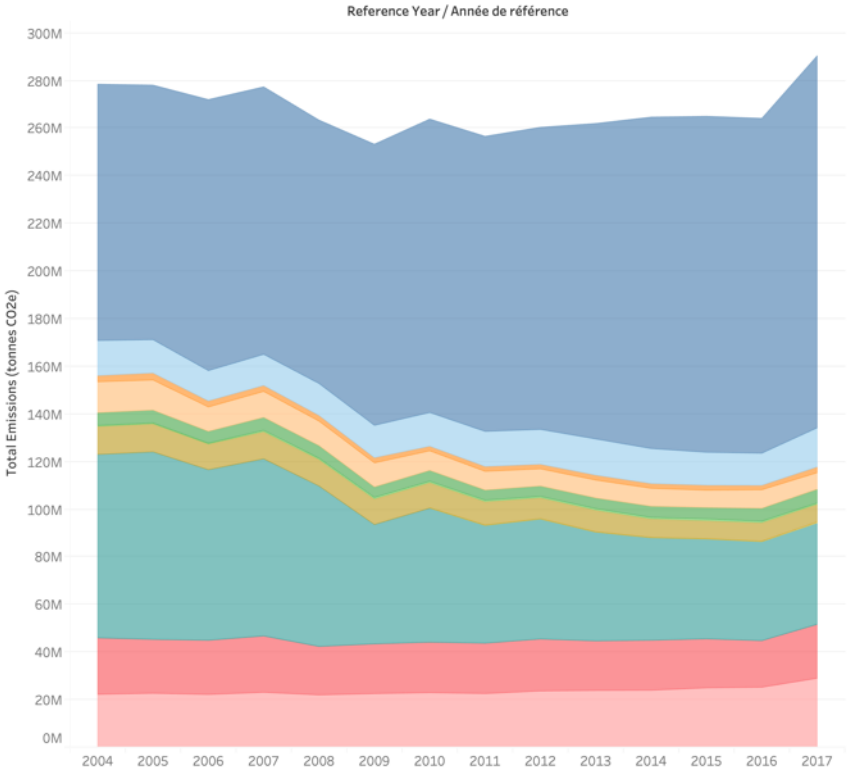

Facility City or District or Municipality / ...
Latitude
Longitude
N2O (tonnes CO2e / tonnes éq. CO2)
N2O (tonnes)
Reporting Company Trade Name / Nom co...
Measure Names
Marks
Add to Sheet
Duplicate
Rename
Hide
Aliases...
Create
Transform
Convert to Measure
Change Data Type
Geographic Role
Default Properties
Group by
Folders
Hierarchy
Replace References...
Describe...
Number (decimal)
Number (whole)
Date & Time
Date
String
Boolean
Default
Measures
CH4 (tonnes CO2e / tonnes éq. CO2)
CH4 (tonnes)
CO2 (tonnes CO2e / tonnes éq. CO2)
CO2 (tonnes)
Reference Year / Année de référence


Facility City or District or Municipality / ...
Latitude
Longitude
N2O (tonnes CO2e / tonnes éq. CO2)
N2O (tonnes)
Reporting Company Trade Name / Nom co...
Measure Names
Marks
Automatic
Add to Sheet
Duplicate
Rename
Hide
Aliases...
Create
Transform
Convert to Continuous
Convert to Measure
Change Data Type
Geographic Role
Default Properties
Group by
Folders
Hierarchy
Replace References...
Describe...
None
Airport
Area Code (U.S.)
CBSA/MSA (U.S.)
City
Congressional District (U.S.)
Country/Region
County
Latitude
Longitude
NUTS Europe
State/Province
ZIP Code/Postcode
Create from
Measures
CH4 (tonnes CO2e / tonnes éq. CO2)
CH4 (tonnes)
CO2 (tonnes CO2e / tonnes éq. CO2)
CO2 (tonnes)
Reference Year / Année de référence
Total Emissions (tonnes CO2e) / Émission...
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values

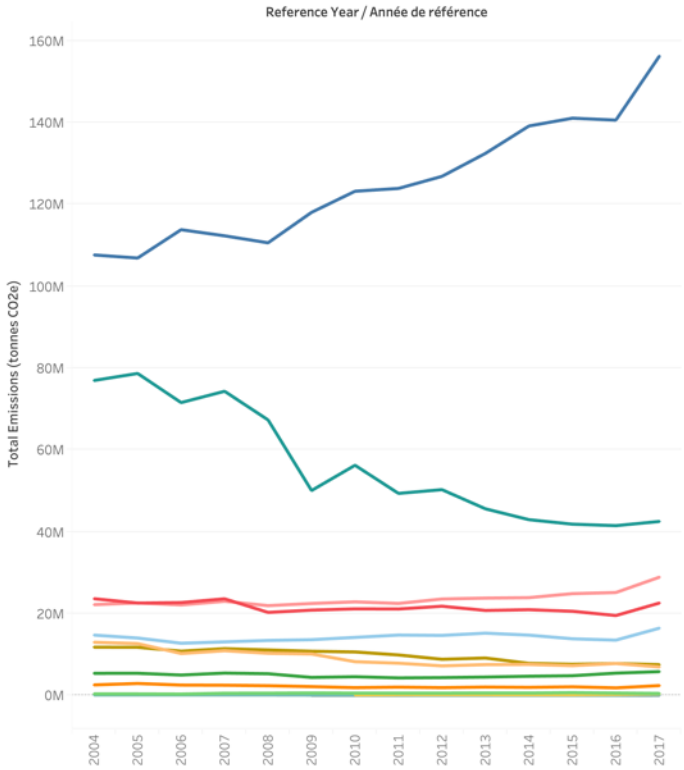

Columns
YEAR(Year)
Rows
SUM(Total Emissions..
GHG by Year
Year
Total Emissions (tonnes CO2e) / Émissions totales (tonnes éq. CO2)
300M
280M
260M
240M
220M
200M
180M
160M
140M
120M
100M
80M
60M
40M
20M
0M
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017


Columns
YEAR(Year)
Rows
SUM(Total Emissions..
GHG by Year
Year
Total Emissions (tonnes CO2e) / Émissions totales (tonnes éq. CO2)
300M
280M
260M
240M
220M
200M
180M
160M
140M
120M
100M
80M
60M
40M
20M
0M
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017

Sheet 1



Sheet 1

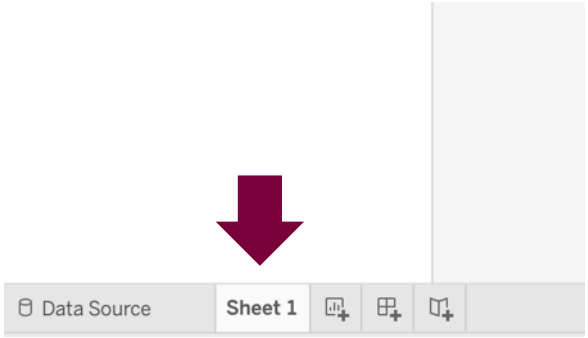


Your turn…

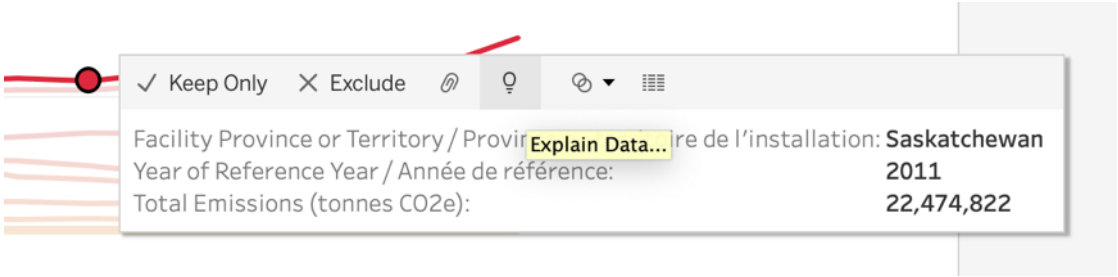
Take a few mins to experiment with:

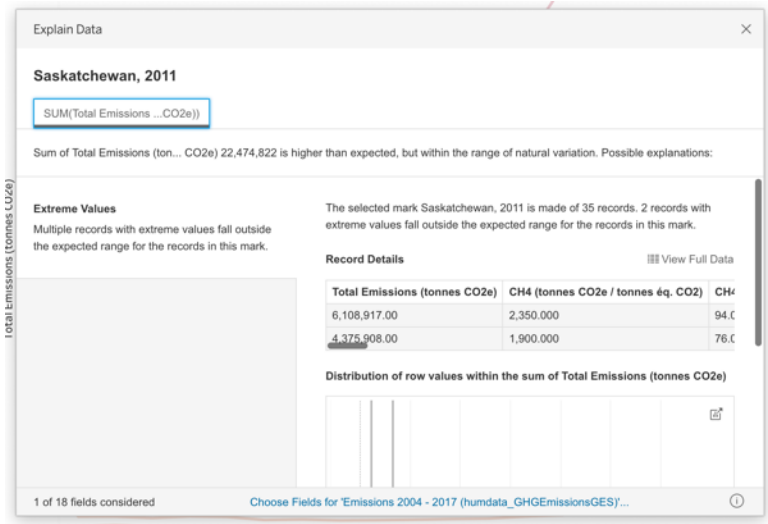
- other measures and dimensions
- different idioms

Does your analysis yield any insights?



Explain data

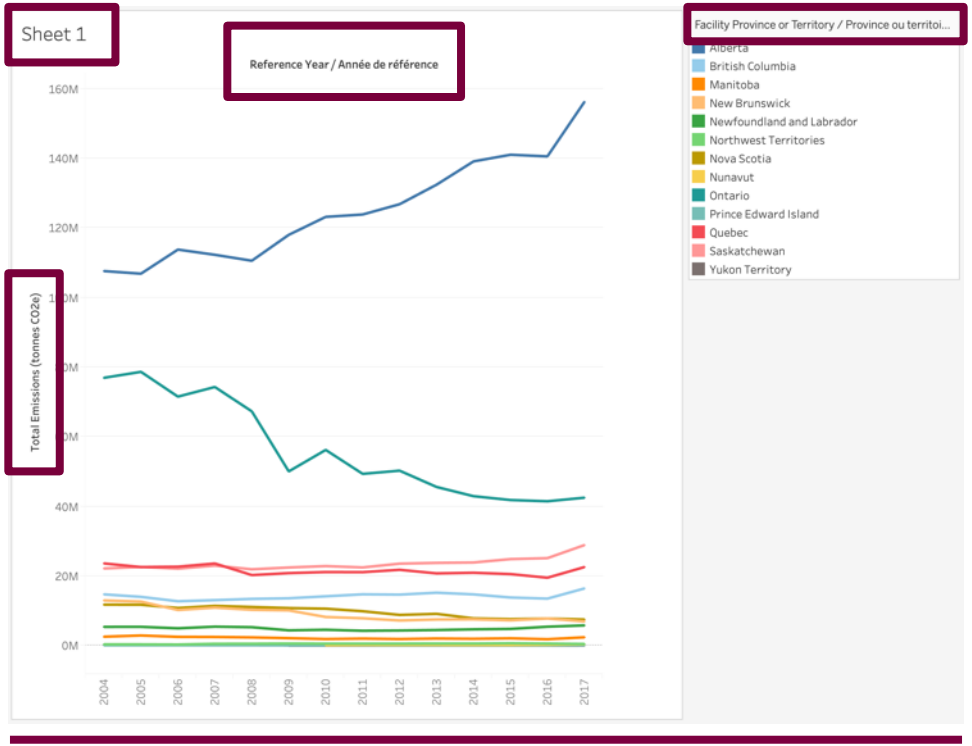




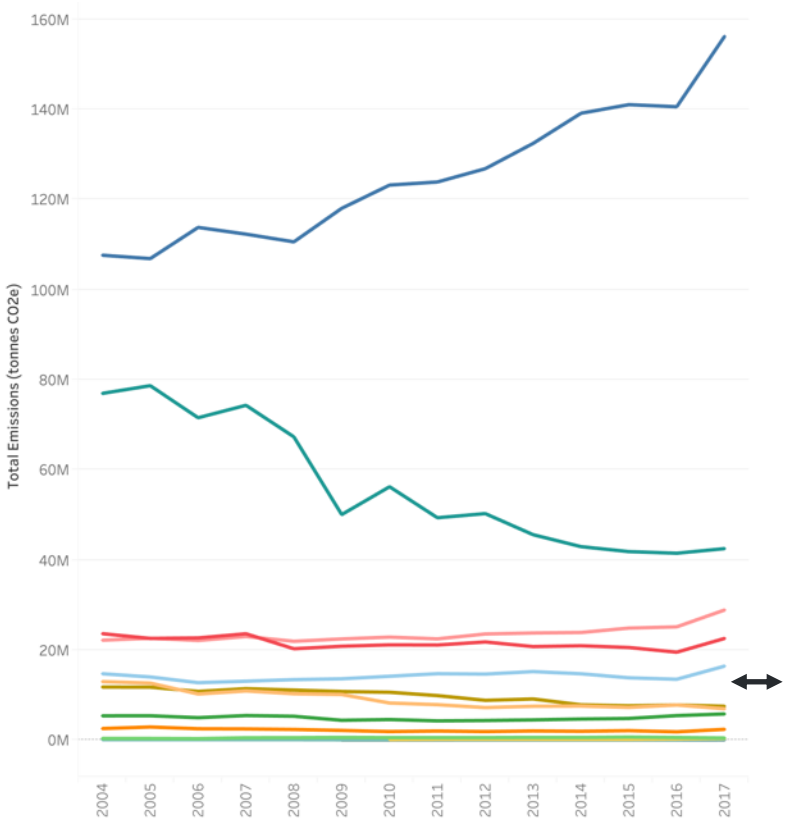
Let’s take a break!

Styling visualization idioms in Tableau

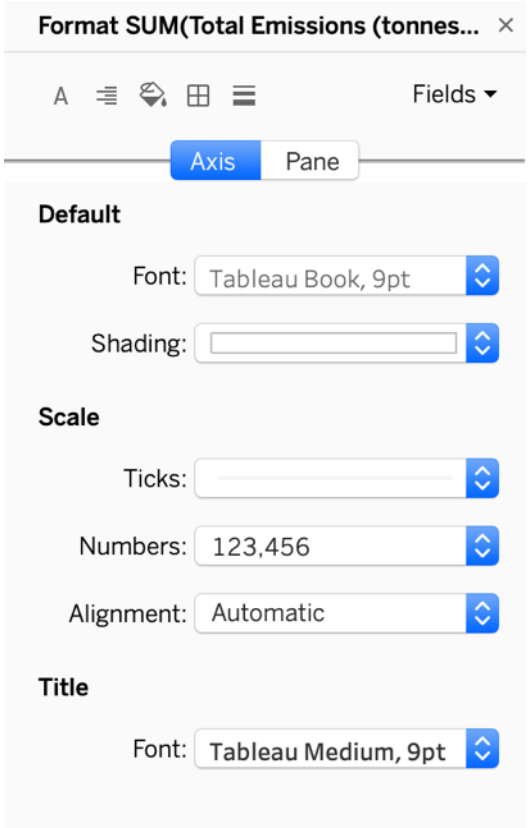
Labels

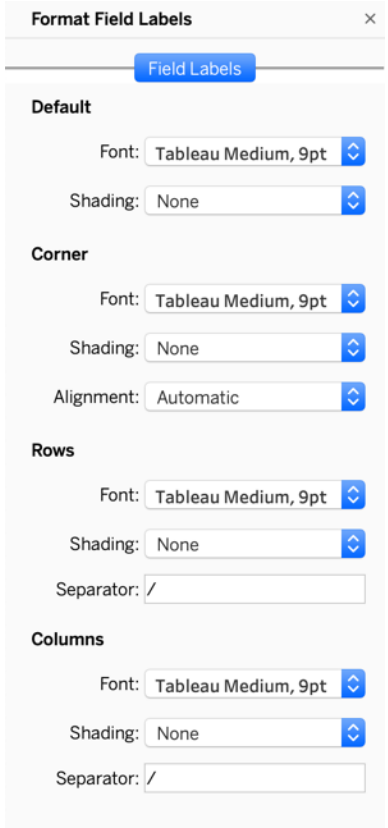

Sheet 1
Reference Year / Année de référence
Total Emissions (tonnes CO2e)
160M
140M
120M
40M
20M
0M
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
Facility Province or Territory / Province ou territoi...
Alberta
British Columbia
Manitoba
New Brunswick
Newfoundland and Labrador
Northwest Territories
Nova Scotia
Nunavut
Ontario
Prince Edward Island
Quebec
Saskatchewan
Yukon Territory

Resize

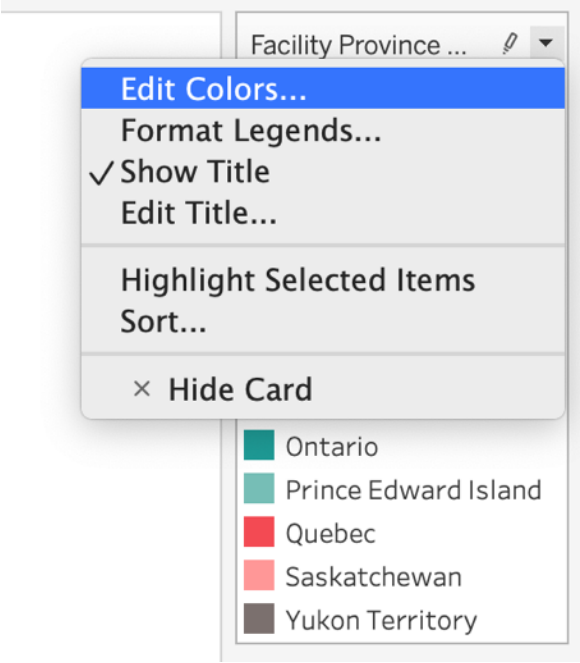

160M
140M
120M
100M
80M
60M
40M
20M
0M
Total Emissions (tonnes CO2e)
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017

Fonts


Format SUM(Total Emissions (tonnes...
Fields
Axis
Pane
Default
Font: Tableau Book, 9pt
Shading:
Scale
Ticks:
Numbers: 123,456
Alignment: Automatic
Title
Font: Tableau Medium, 9pt

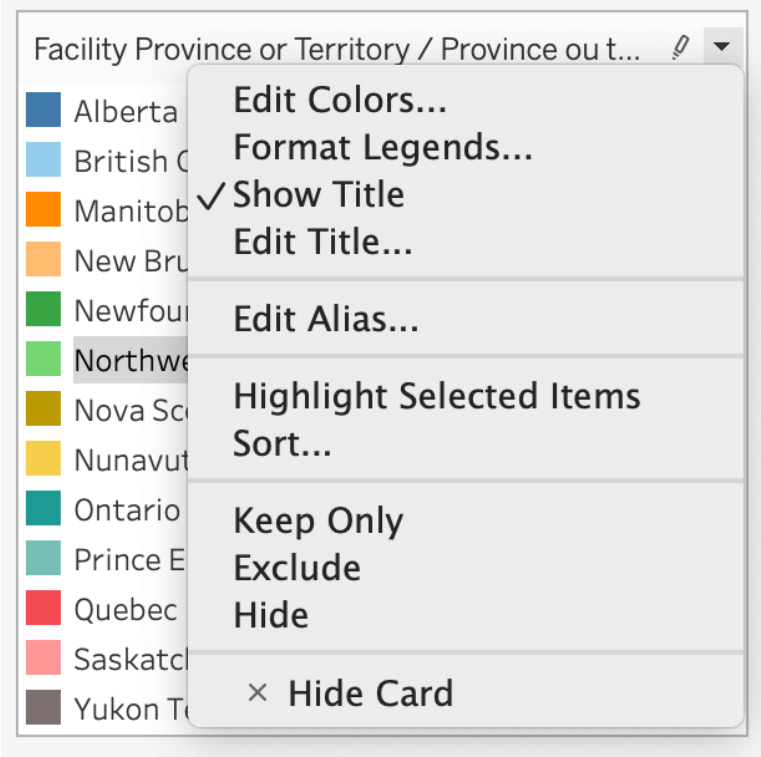

Format Field Labels
Field Labels
Default
Font: Tableau Medium, 9pt
Shading: None
Corner
Font: Tableau Medium, 9pt
Shading: None
Alignment: Automatic
Rows
Font: Tableau Medium, 9pt
Shading: None
Separator: /
Columns
Font: Tableau Medium, 9pt
Shading: None
Separator: /

Colours


Facility Province ...
Edit Colors...
Format Legends...
✓ Show Title
Edit Title...
Highlight Selected Items
Sort...
× Hide Card
Ontario
Prince Edward Island
Quebec
Saskatchewan
Yukon Territory

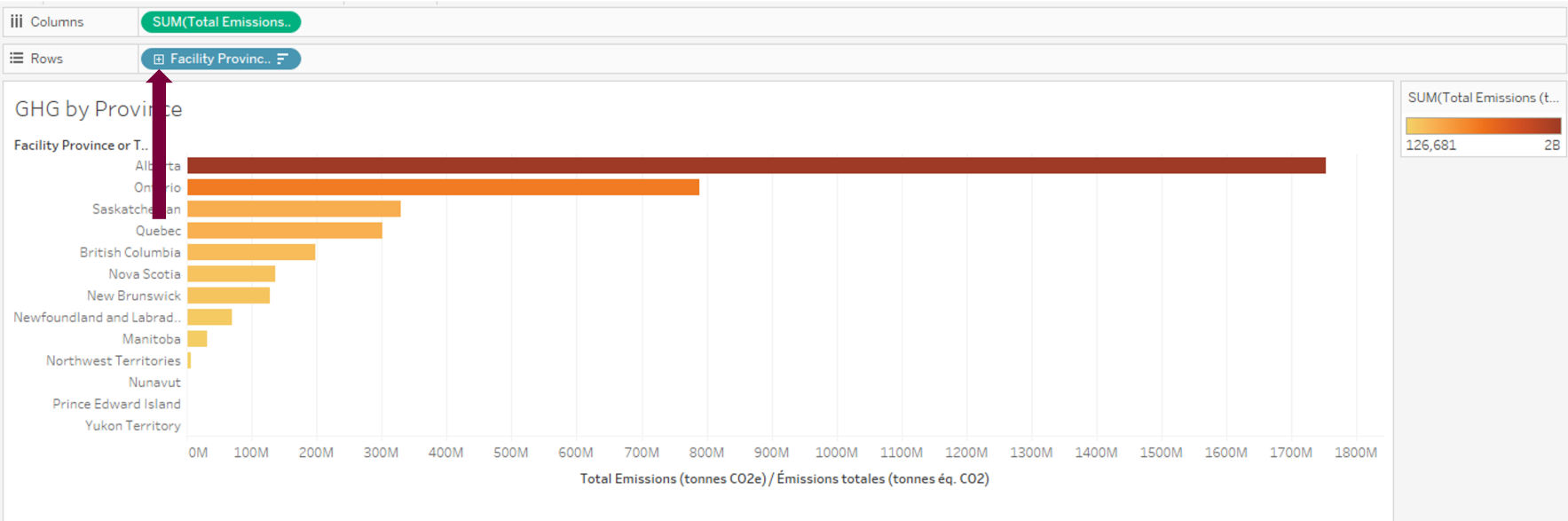
Edit Colors [Facility Province or Territory / Province ou territoire de l'installation]
Select Data Item:
Alberta
British Columbia
Manitoba
New Brunswick
Newfoundland and Labrador
Northwest Territories
Nova Scotia
Nunavut
Ontario
Prince Edward Island
Quebec
Saskatchewan
Yukon Territory
Select Color Palette:
Automatic
Tableau 10
Tableau 20
Color Blind
Seattle Grays
Traffic Light
Superfishel Stone
Miller Stone
Nuriel Stone
Jewel Bright
Summer
Winter
Green-Orange-Teal
Blue-Red-Brown
Purple-Pink-Gray
Tableau Classic 10
Tableau Classic Medium
Tableau Classic 20
Blue
Orange
Green
Red
Purple
Brown
Gray
Gray Warm
Blue-Teal
Orange-Gold
Green-Gold
Red-Gold
Hue Circle
Reset
Apply

Legend

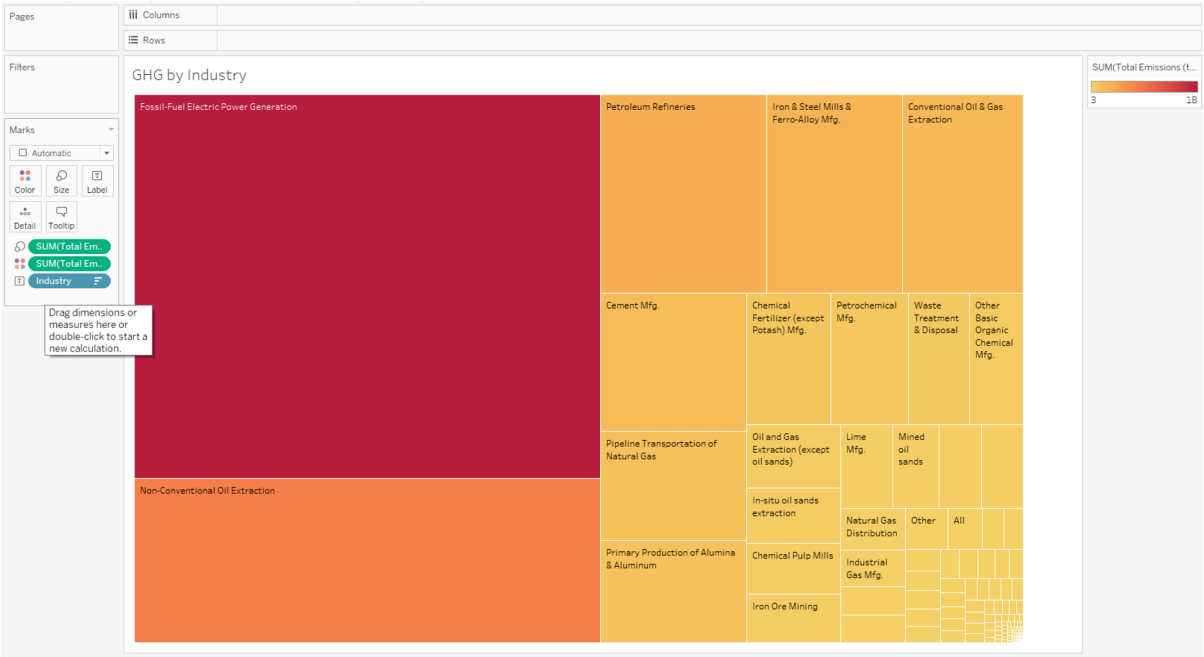

Facility Province or Territory / Province ou t...
Alberta
British C
Manitob
New Bru
Newfoul
Northwe
Nova Sc
Nunavut
Ontario
Prince E
Quebec
Saskatc
Yukon T
Edit Colors...
Format Legends...
Show Title
Edit Title...
Edit Alias...
Highlight Selected Items
Sort...
Keep Only
Exclude
Hide
Hide Card

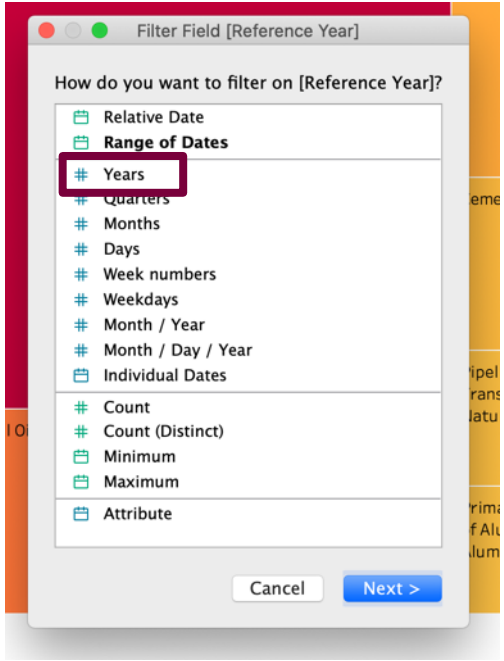
Building a dashboard in Tableau

Viz 2: GHG emissions by province

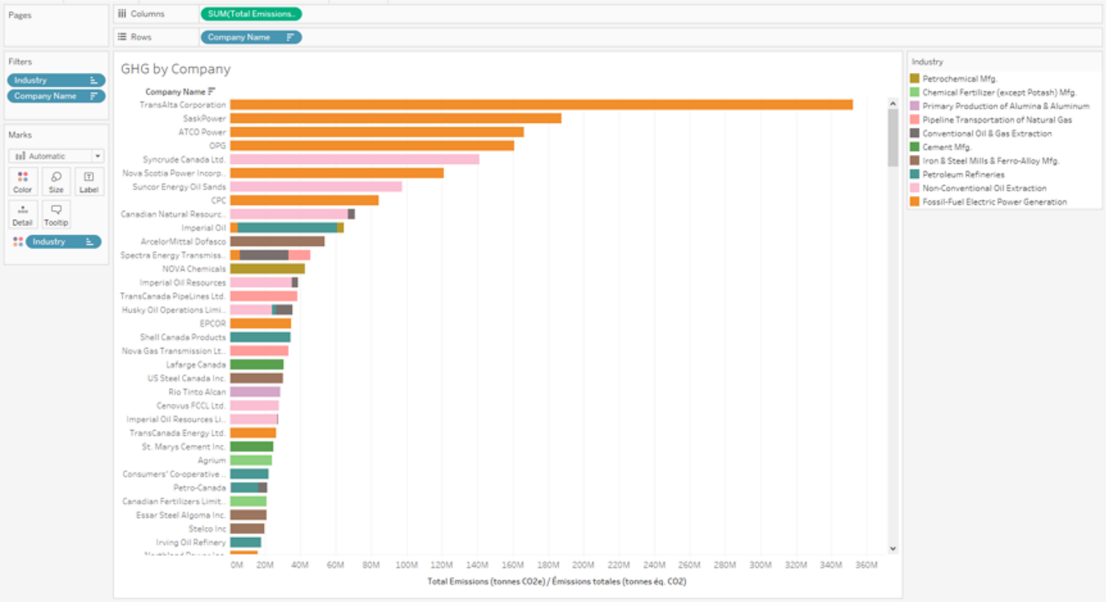


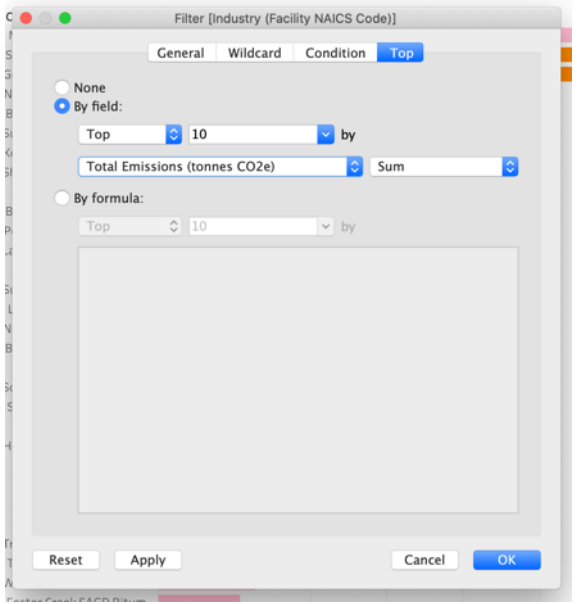
Viz 3: GHG emissions by industry



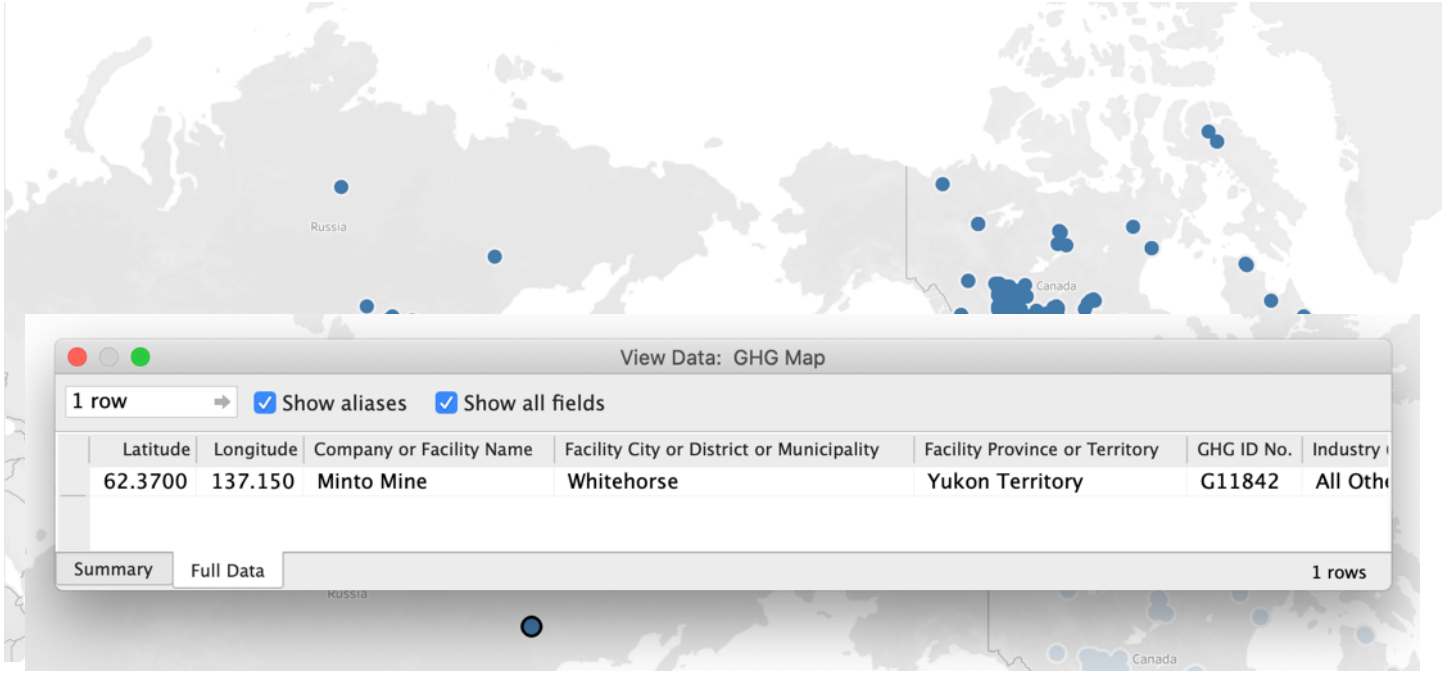


Viz 4: GHG emissions by company

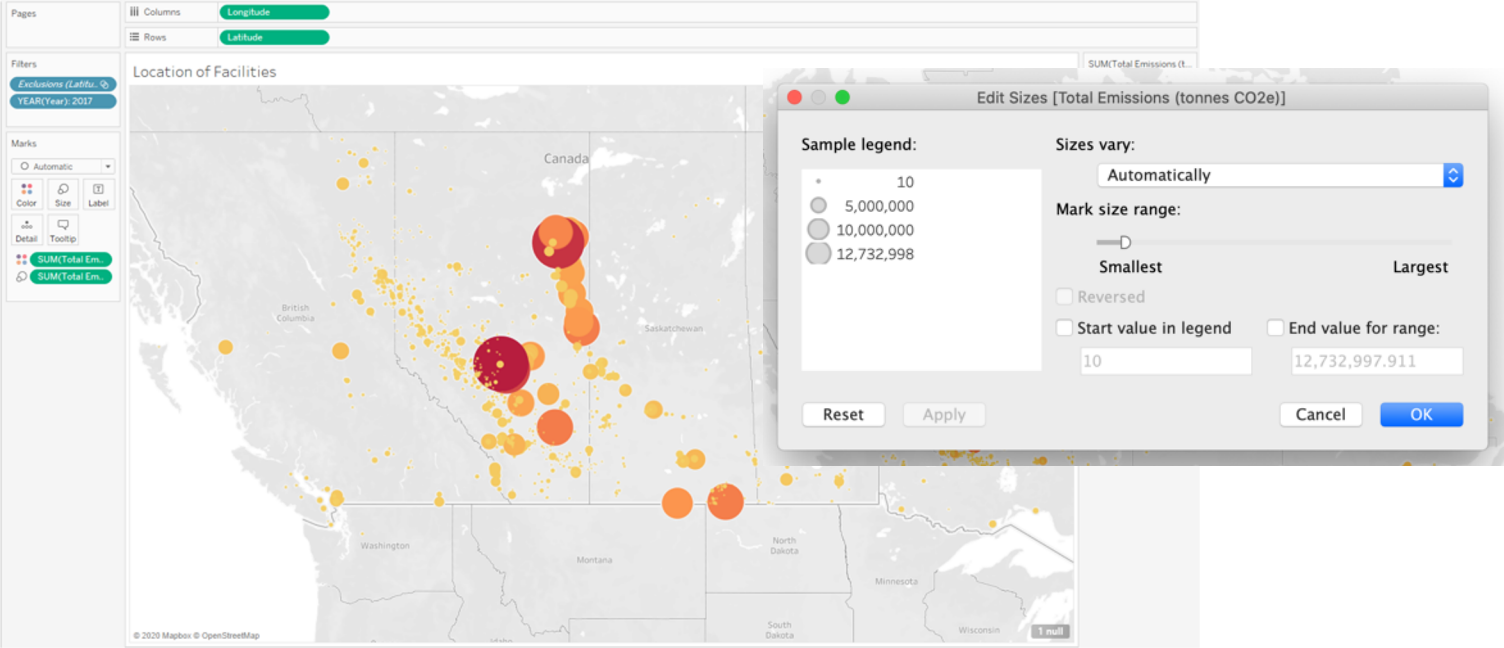




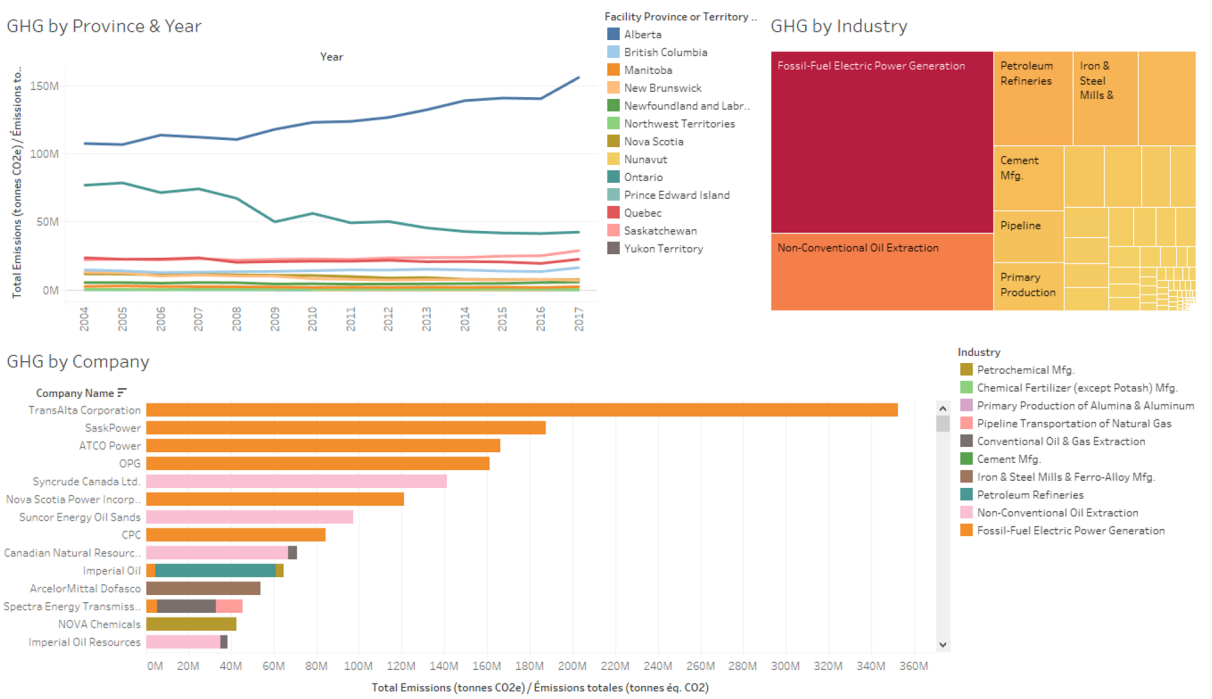
Viz 5: Mapping latitude and longitude – what's up with the data?



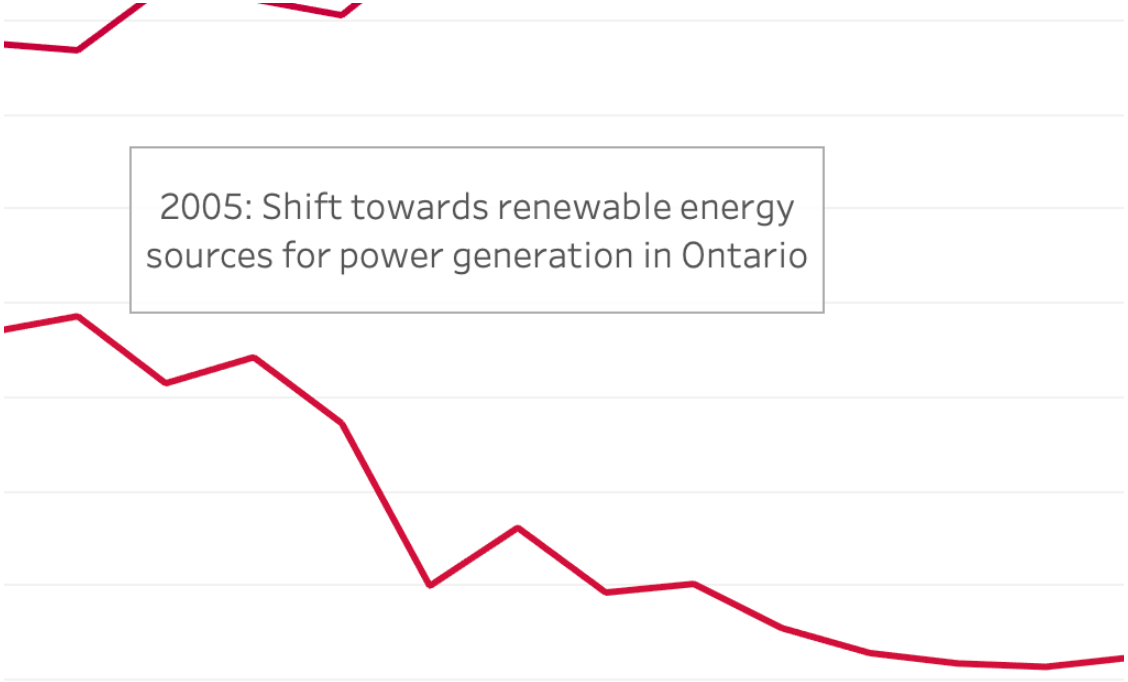
Viz 5: Mapping latitude and longitude – what's up with the data?



Create a dashboard to bring multiple visualizations together...



Stories allow you to contextualize the data



Visualization provenance

- I.e. documenting what steps were involved in creating the visualization
- For you: for tracing errors and reproducing steps
- For others: for evaluating the quality of the data and the validity of the visualization output

More resources

Tableau is waaaaay more powerful than what we've explored today...

→ take your learning further with Tableau articles and videos.

Also: Munzner, T. *Visualization Analysis and Design*. CRC Press, 2015.

Thank you for coming!

Feel free to get in touch with questions!

mordelld@mcmaster.ca