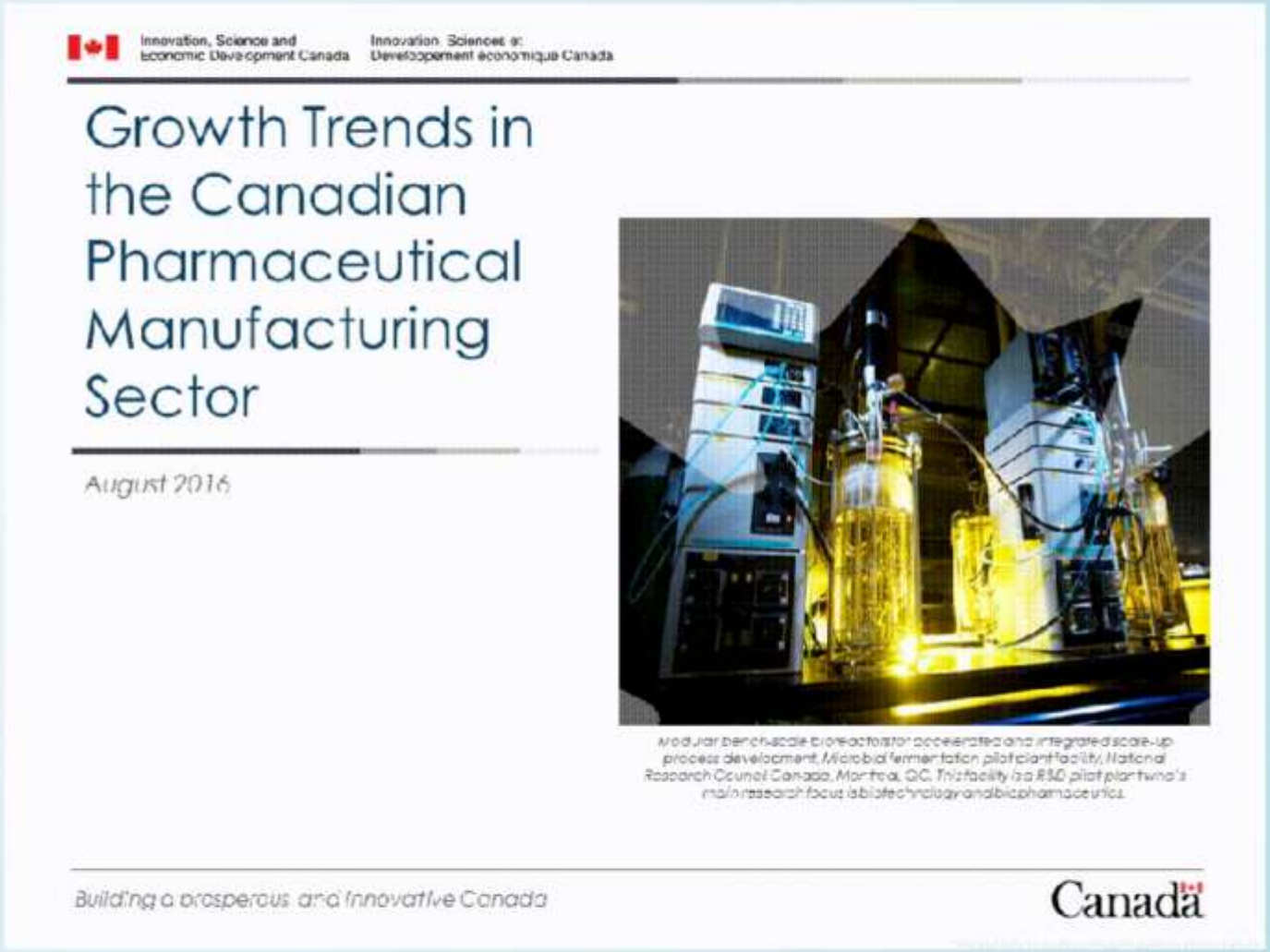
Automating industry sector trend analysis using R and Statistics Canada open data API/Web data services

By Jacob Raymond, MLSB-LSID

Project Genesis

NAICS 3254 CANSIM Data Extract (Aug. 2016)

- Trends deck focused on NAICS 3254.
- Included plots and economic analysis.



Innovation, Science and Economic Development Canada | Innovation, Sciences et Développement économique Canada

Growth Trends in the Canadian Pharmaceutical Manufacturing Sector

August 2016

Modular benchscale bioreactors for accelerated and integrated scale-up process development, Microbial fermentation pilot plant facility, National Research Council Canada, Montreal, QC. This facility is a R&D pilot plant which's main research focus is biotechnology and biopharmaceutics.

Building a prosperous and innovative Canada

Canada

Project Rationale

- Data extraction and interpretation is crucial to understanding sectoral needs and opportunities, and for evidence-based policy making.

- ISED has accessible to multiple internal databases and tools.

- Statistics Canada's data portal (formally CANSIM) is a useful, open-access resource in this respect.
  - However, it can quickly become overwhelming.

Proposed Solution

An algorithmic tool that outputs up-to-date results into a visualization-based deck.

Slide Example

Gross Domestic Product

In October 2018, food and beverage manufacturing accounted for $34780M of gross domestic product, representing 17.18% of the manufacturing gross domestic product.

Gross domestic product (GDP) at basic prices, by industry, monthly. Vector IDs = v65201210, v65201263, v65201264, v65201275, v65201276, v65201277, v65201280

Automatically generated sentence

Plot always incorporates latest data

Footer generated automatically

How it Works

1. Import the list of vectors
2. For every variable of interest:
    1. For every vector:import observations for latest N years (using API). Save in table
    2. If there's a group:
        - sum (if index series)
        - take the mean (if unindexed)
    3. If series is indexed: index
    4. Save metadata
    5. Create a plot with the data
3. Use the plots and metadata to create a deck

What is R?

R is an object-based programming language. It is similar to Python, except that it's geared towards statistical analysis and data visualization.

R is free to download and does not require administrator permission.

Source of Image: Hadley Wickham and others at RStudio - https://www.rproject.org/logo/, CC BY-SA 4.0

What is the Web Data Service (WDS)?

Basically, it's a collection of tools to extract data directly from Statistics Canada's server. There are currently 12 methods available.

```
#Load some packages
pacman::p_load(httr)

#Configure the proxy (IMPORTANT, else this won't work)
set_config(use_proxy(url="cdhwg01.prod.prv",port=80))


#Suppose we want to know the probation rate per 10,000 young persons for Newfoun

#The first step is to find the corresponding vector. The only way to get it for
#To do this, use the getFullTableDownloadCSV method.

pid<-"27100333" #Enter table number (In this case 35100003)

#This method basically returns the url to download the full table (in a zipped f
```

Some Definitions

- Vector: A unique alphanumeric code that refers to a particular time series.
- Product Identification Number (PID): A unique number assigned to every Statistics Canada product.

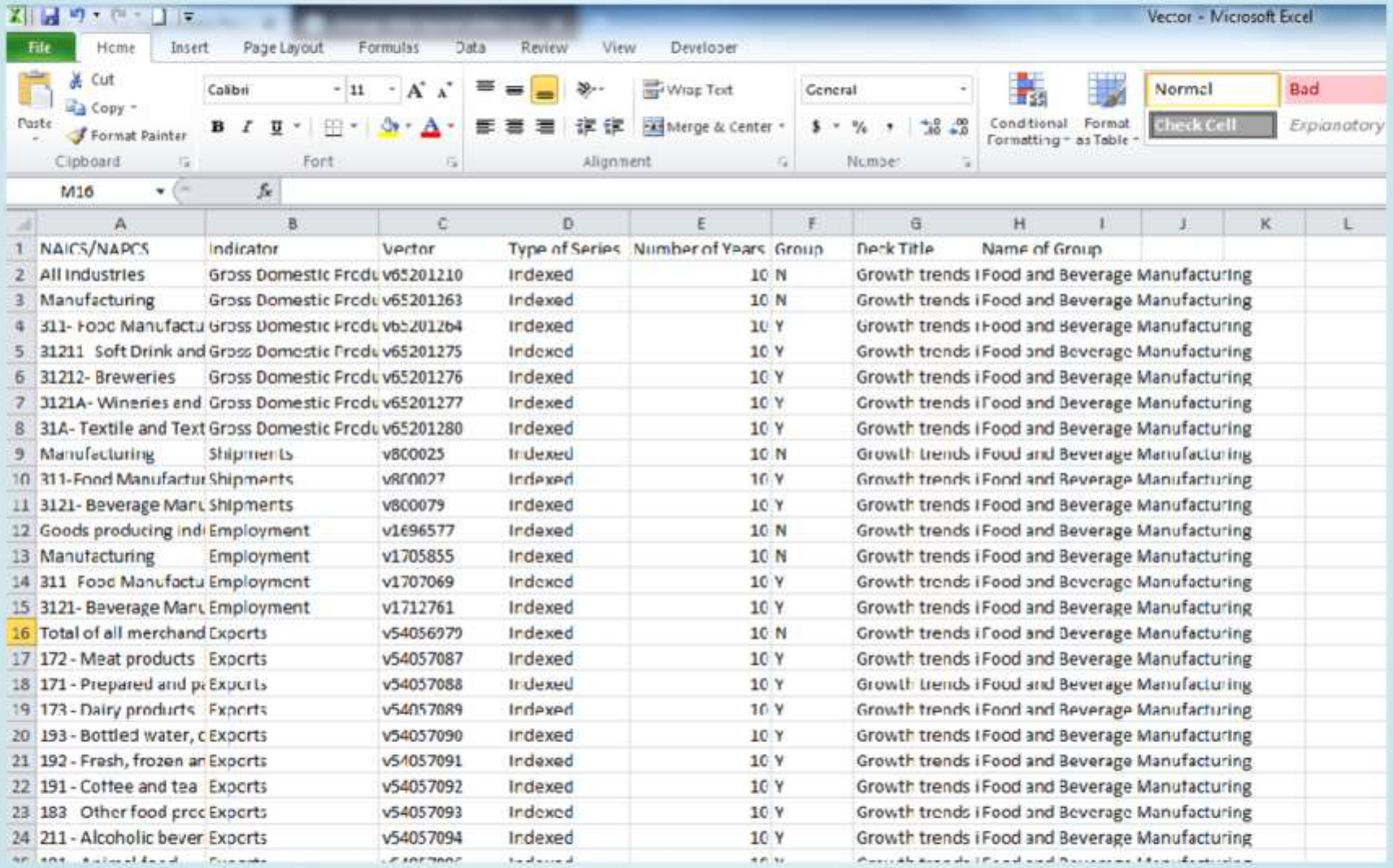
WDS Methods Used

- getSeriesInfoFromVector: Obtain series metadata from vector.

- getCubeMetadata: Obtain table metadata from pid.

- getDataFromVectorsAndLatestNPeriods: Extract vector data for the latest N periods.

List of Vectors



Vector - Microsoft Excel

File | Home | Insert | Page Layout | Formulas | Data | Review | View | Developer

Cut | Copy | Format Painter | Paste | Clipboard | Calibri | 11 | Font | Wrap Text | Merge & Center | Alignment | General | Number | Conditional Formatting | Format as Table | Normal | Bad | Check Cell | Explanatory

M16

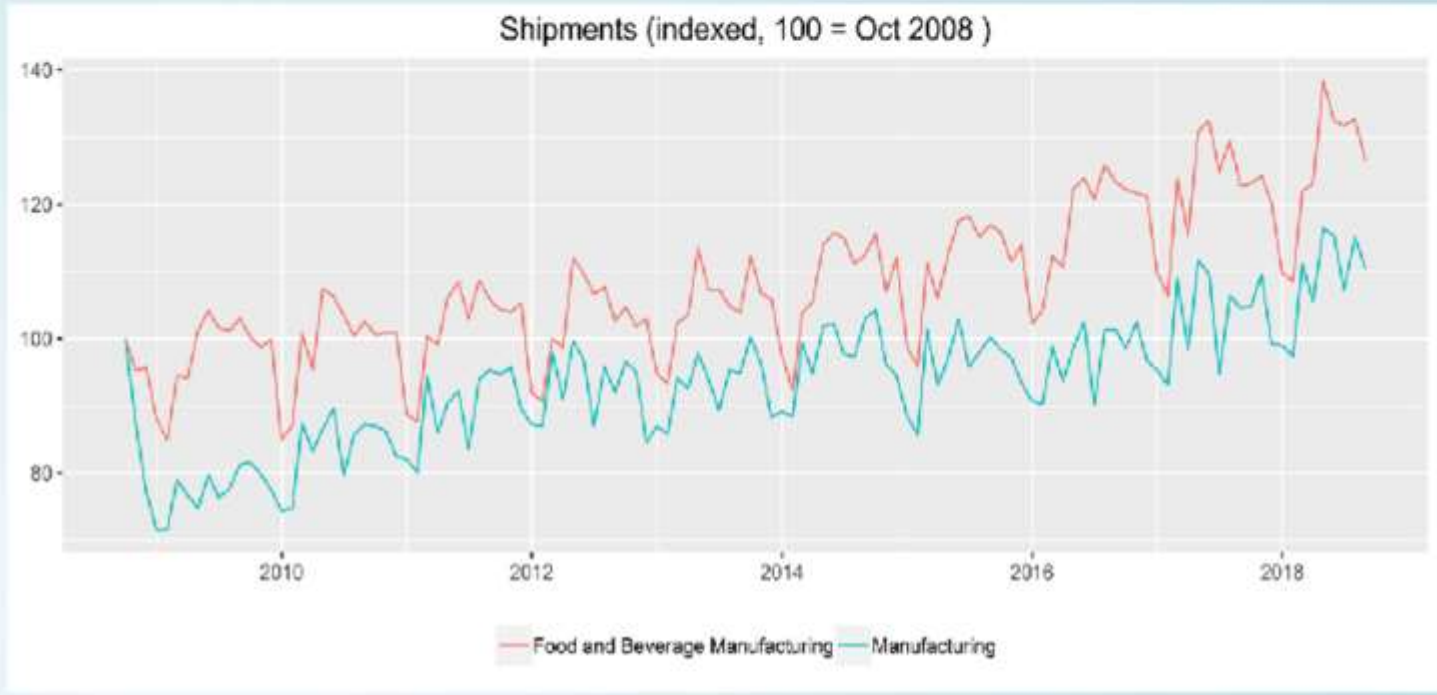
	A	B	C	D	E	F	G	H	I	J	K	L
1	NAICS/NAPCS	Indicator	Vector	Type of Series	Number of Years	Group	Desk Title	Name of Group				
2	All Industries	Gross Domestic Produ	v65201210	Indexed	10	N	Growth trends i	Food and Beverage Manufacturing				
3	Manufacturing	Gross Domestic Produ	v65201263	Indexed	10	N	Growth trends i	Food and Beverage Manufacturing				
4	311- Food Manufactu	Gross Domestic Produ	v65201264	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
5	31211 Soft Drink and	Gross Domestic Produ	v65201275	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
6	31212- Breweries	Gross Domestic Produ	v65201276	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
7	3121A- Wineries and	Gross Domestic Produ	v65201277	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
8	31A- Textile and Text	Gross Domestic Produ	v65201280	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
9	Manufacturing	Shipments	v800025	Indexed	10	N	Growth trends i	Food and Beverage Manufacturing				
10	311-Food Manufactur	Shipments	v800027	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
11	3121- Beverage Manu	Shipments	v800079	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
12	Goods producing ind	Employment	v1696577	Indexed	10	N	Growth trends i	Food and Beverage Manufacturing				
13	Manufacturing	Employment	v1705855	Indexed	10	N	Growth trends i	Food and Beverage Manufacturing				
14	311 Food Manufactu	Employment	v1707069	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
15	3121- Beverage Manu	Employment	v1712761	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
16	Total of all merchand	Exports	v54056979	Indexed	10	N	Growth trends i	Food and Beverage Manufacturing				
17	172 - Meat products	Exports	v54057087	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
18	171 - Prepared and pe	Exports	v54057088	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
19	173 - Dairy products	Exports	v54057089	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
20	193 - Bottled water, c	Exports	v54057090	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
21	192 - Fresh, frozen an	Exports	v54057091	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
22	191 - Coffee and tea	Exports	v54057092	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
23	183 Other food proc	Exports	v54057093	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				
24	211 - Alcoholic bever	Exports	v54057094	Indexed	10	Y	Growth trends i	Food and Beverage Manufacturing				

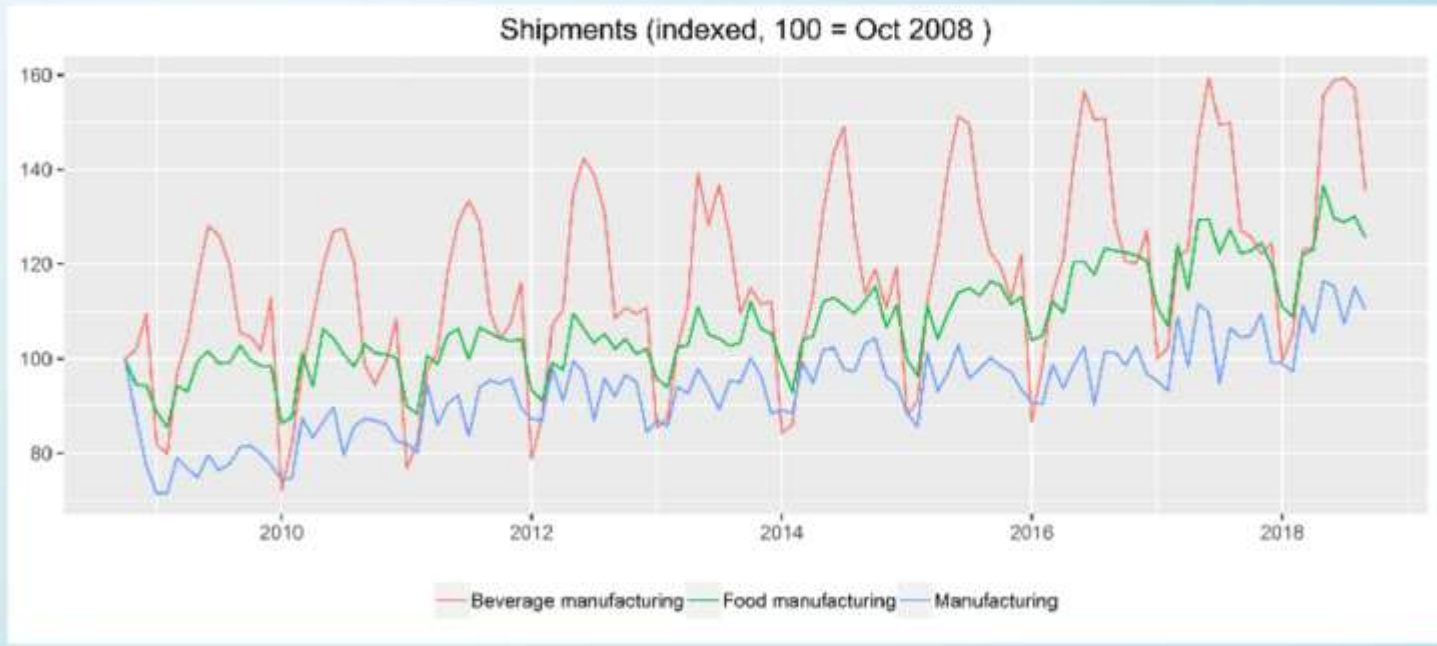
User submitted narrow data saved as a CSV file

Grouping

Grouping is useful if you want to represent a category of NAICS/NAPCS.

- If series is indexed: sum, then index
- If unindexed: mean





An Example: Food and Beverage Manufacturing

Some Possible Next Steps

Technical

- Visual appeal
- Troubleshooting
- Weighted average for unindexed series
- Save functions in package
- Share code on GitHub, Wiki, etc.
- Improve template usability

Some Possible Next Steps

Long Term Vision

- Forecasting

- Make the tool accessible to stakeholders and,ultimately, the general public

- Detect if time series should be indexed or not

- Improve the language in the descriptive sentence, make it seem more human

Lessons Learned

- R as a new professional paradigm.
- Adapting to Statistics Canada's data storage and presentation mechanisms.
- Some features were significantly easier to implement than first expected.
- Automation is the way to go.

Questions? Comments?