Data Explorer 2.0: Analyse and Reuse data in your repository

Dataverse Workshop, April 17, 2023

Borealis (https://borealisdata.ca)

———

Canadian Dataverse Repository, provided in partnership with Canadian academic libraries.

- Bilingual platform
- Institutional Collections
- Open repository development
- Shared standards

Data exploration

- Helps users understand the data
  - Preview data, documentation, codebooks, readme
  - Get started with data; digging, filtering, previewing
- Identify and find data
  - Variables and questions
  - Frequencies, statistics, cross-tabulation
  - Chart visualization
  - Subsetting & downloading
- Analyse and reuse training
  - Transparency and reproducibility
  - Teaching data concepts
  - Open access and sharing
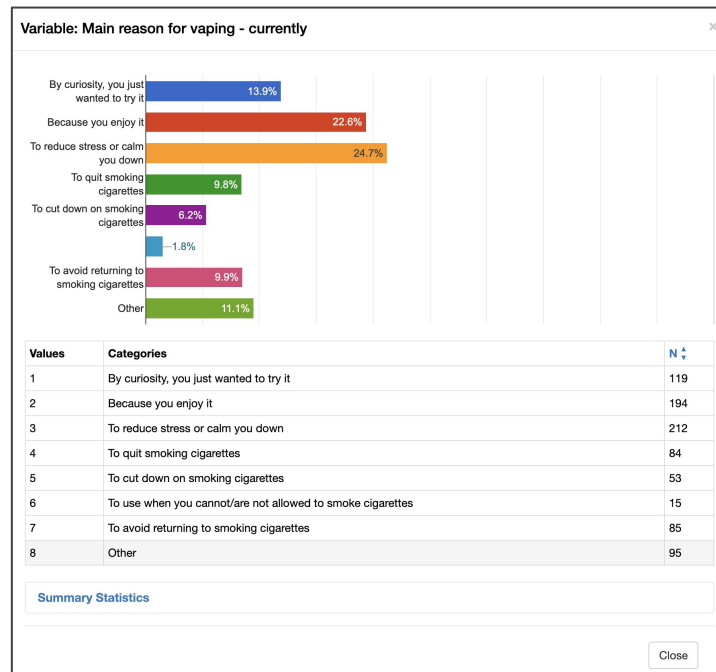
Source: VAP_35R, Canadian Tobacco and Nicotine Survey, 2021.
Odesi https://odesi.ca

Dataverse Repository

———

- Social science datasets are supported in Dataverse (many features)
  - Descriptive metadata
  - Variable metadata
  - Tabular data ingest (TAB file)
  - File level APIs
- Metadata Standards
  - Data Documentation Initiative (DDI)
  - Dataverse JSON
  - Other disciplinary connections
- Dataverse APIs
- R Tools
- Integrates with Data Explorer (1.0, 2.0)

Data Explorer

———

- Dataverse Tabular file (.tab) & Data Explorer
- Preview tabular data
  - Variables & frequencies
- Summary statistics
  - (Averages, category frequencies, tabulations)
- Integrated external application since Dataverse version 4.6
- Data Explorer 1.0-2.0 in [Github](#)
  - Angular web framework
  - New in 2.0:
    - Variable Groups
    - View Question Text, Universes
    - Frequencies now include weights
    - Updated cross-tabulation feature
      - Multivariate analysis

Canadian Tobacco and Nicotine Survey, 2021

CTNS-5305-E-2021_F1.tab

English ▾

‹ Hide Groups

⊞ Cross Tabulation   ≡ Summary Statistics   Download ▾

All Variables
CAN: Cannabis
DV: Derived Variables
ALC: Alcohol
IU: Initial use
VAP: Vaping
OTP: Other tobacco product status
RD: Reference date
DEM2: Demographics 2
SRV: Survey Related Variables
GDR: Gender
TBC: Tobacco

Search 🔍

Items per page 25 ▾   1 - 25 of 104   ‹   ›

	ID	Name	Label	Weight	View Summary Statistics	View Categories	View Questions
☐	v6423547	AGEGROUP	Age group of person		▤	▥	💬
☐	v6423552	ALC_05	Frequency drank alcoholic beverage - past 30 days		▤	▥	💬
☐	v6423564	ALC_10	Freq drank 4 drinks or more on one occasion - past 12 months		▤	▥	💬
☐	v6423571	CAN_05AR	Smoked cannabis - lifetime		▤	▥	💬
☐	v6423570	CAN_05BR	Age first time smoked cannabis		▤	▥	💬
☐	v6423525	CAN_10AR	Frequency smoked cannabis - past 30 days		▤	▥	💬
☐	v6423539	CAN_10BR	Nb of days smoked cannabis at least once past month - grouped		▤	▥	💬
☐	v6423524	CAN_15AR	Frequency smoked cannabis/tobacco mix - past 30 days		▤	▥	💬
☐	v6423569	CAN_15BR	Nb of days smoked can./toba. mix at least once a week - past 30 days		▤	▥	💬
☐	v6423529	CAN_17R	Frequency consumed edibles - past 30 days		▤	▥	💬
☐	v6423486	CAN_20AR	Vaped cannabis - lifetime		▤	▥	💬
☐	v6423483	CAN_20BR	Age first time vaped cannabis - grouped		▤	▥	💬
☐	v6423541	CAN_25AR	Frequency vaped cannabis - past 30 days		▤	▥	💬

Data Explorer tools

Depositing and submitting data

———

- Upload tabular files (SPSS, STATA, SAS, R, CSV, Excel)
- Data ingest tips:
  - Responses/ observations as rows
  - Variables/concepts as columns
  - No notes / improper formatting
  - Does not support multiple spreadsheets in one file
  - Excel (XLST) is supported
  - CSV requires documentation
- Guides and other resources

Variable metadata

___

- Assigns Unique identifiers
- Names, labels, codes, values
- Description of content
  - Interviews, question text, universes, recodes and notes
- Weighting
- Summary statistics
- Export HTML and PDF Codebooks
- Export XML metadata for integrations

Source: Canadian Tobacco and Nicotine Survey, 2021.
DDI Codebook (Odesi)

VAP_05BR: Age first time tried vaping
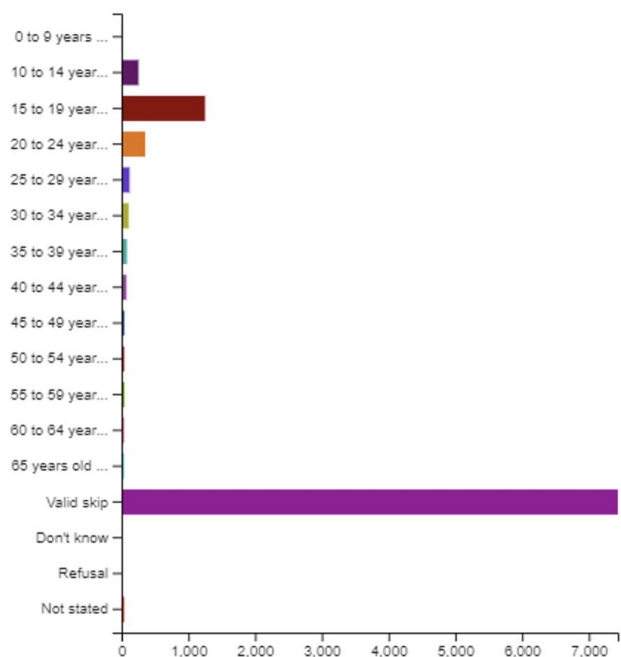
Values	Categories	Count	Count Percentage(%)	Weighted Count	Weighted Count Percentage(%)
1	0 to 9 years old	5	0.05		
2	10 to 14 years old	253	2.553		

Variable Information

ID: v6423478

Name: VAP_05BR

Label: Age first time tried vaping

Literal Question: How old were you when you first tried vaping?

Universe: VAP_05AR = 1

Weight Variable: Unweighted

Is Weight: No

Editing variable metadata

———

- Variable metadata is ingested and stored in Dataverse database
- Related [Data Curation Tool](#) supports editing variable metadata
  - Add variable groups, question/concept description, universe notes, apply weights
- Edits are saved as new version in DV
- Updates are displayed in Data Explorer

Connecting to Data Explorer

Preview data

— — —

smoking policies on smoking behaviour, the Canadian Tobacco Use Monitoring

EXPLORE VARIABLES ↗	MATCHING VARIABLES 96

ID	Name	Label
1	v2609851	Any in hhld smoke cigarettes
2	v2609786	# of people that smoke inside the home
3	v2609819	# of cigs. smoked inside home
4	v2609923	Is smoking cigs. allowed in your home
5	v2609850	Is smoking cigs. restricted in your home
6	v2609818	Smoking restricted-in certain rooms only
7	v2609814	Smoking restricted-presence of children
8	v2609808	Smoking restricted-windows open
9	v2609789	Smoking restricted-other
10	v2609815	Smoke cigs. daily/occasional/not at all
11	v2609877	Past 30 days, smoked any cigarettes
12	v2609853	Past 30 days, smoke every day
13	v2609874	Smoked at least 100 cigarettes in life
14	v2609898	Smoked a whole cigarette
15	v2609845	Smoked cigarettes daily
16	v2609839	Age smoked first cigarette

Variable: Frequency smoked cigarettes - past 30 days ✕
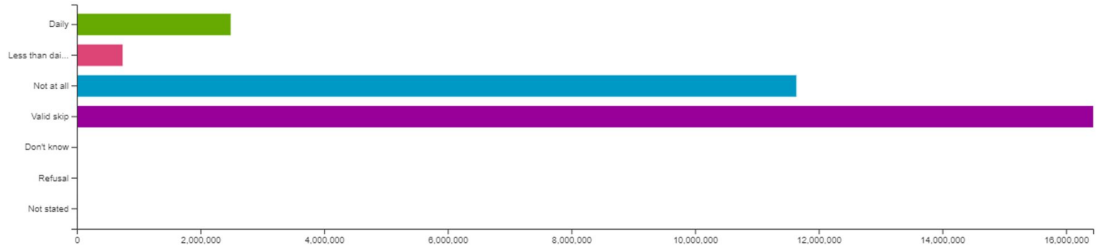
Values	Categories		Count	Count Percentage(%)	Weight Count
1	Daily		608	7.495	2487936.268
2	Less than daily, but at least once in the past month		184	2.268	740806.37
4	Not at all		2852	35.158	1.1637711838E7
6	Valid skip		4464	55.030	1.6439278195E7
7	Don't know		0	0.000	0.0
8	Refusal		0	0.000	0.0
9	Not stated		4	0.049	890.84

Analyse data

Select Variables For Cross Tabulation

VAP_15AR - Nb of days vaped e-liquid with nicotine - pas ☐

VAP_15BR - Nb of days vaped e-liquid without nicotine - ☐

VAP_15CR - Nb of days vaped e-liquid didn⬚t know what it ☐

VAP_20R - Nb of times pick up vaping device - Day respo ☐

VAP_21R - Nb of puffs taken each time vaping device is ☐

VAP_30R - Flavour respondent vape most often ☐

VAP_40AR - Where get vaping devices - Vape shop ☐

VAP_40BR - Where get vaping devices - Convenience store/ ☐

VAP_40CR - Where get vaping devices - Supermarket/grocer ☐

VAP_40DR - Where get vaping devices - Online ☐

→

Rows

VAP_35R - Main reason for vaping - currently ☐

→

Columns

AGEGROUP - Age group of person ☐

OK

Cross Tabulation Table

AGEGROUP - Age group of person	15 to 19 years old	20 to 24 years old	25 to 34 years old	35 to 44 years old	45 to 54 years old	55 to 64 years old	65 years old and older	Valid skip	Don't know	Refusal	Not stated	Total
VAP_35R - Main reason for vaping - currently												
By curiosity, you just wanted to try it	59 0.6%	45 0.45%	7 0.07%	3 0.03%	3 0.03%	1 0.01%	1 0.01%	0 0%	0 0%	0 0%	0 0%	119 1.2%
Because you enjoy it	72 0.73%	102 1.03%	9 0.09%	7 0.07%	2 0.02%	1 0.01%	1 0.01%	0 0%	0 0%	0 0%	0 0%	194 1.96%
To reduce stress or calm you down	87 0.88%	105 1.06%	7 0.07%	8 0.08%	2 0.02%	0 0%	3 0.03%	0 0%	0 0%	0 0%	0 0%	212 2.14%
To quit smoking cigarettes	6 0.06%	37 0.37%	10 0.1%	11 0.11%	12 0.12%	2 0.02%	6 0.06%	0 0%	0 0%	0 0%	0 0%	84 0.85%
To cut down on smoking cigarettes	1 0.01%	28 0.28%	4 0.04%	8 0.08%	3 0.03%	5 0.05%	4 0.04%	0 0%	0 0%	0 0%	0 0%	53 0.53%
To use when you cannot/are not allowed to smoke cigarettes	2 0.02%	8 0.08%	0 0%	2 0.02%	2 0.02%	1 0.01%	0 0%	0 0%	0 0%	0 0%	0 0%	15 0.15%
To avoid returning to smoking cigarettes	2 0.02%	41 0.41%	15 0.15%	8 0.08%	12 0.12%	3 0.03%	4 0.04%	0 0%	0 0%	0 0%	0 0%	85 0.86%
Other	31 0.31%	49 0.49%	3 0.03%	6 0.06%	2 0.02%	2 0.02%	2 0.02%	0 0%	0 0%	0 0%	0 0%	95 0.96%
	1775	1987	641	912	841	1109	1764	0	0	0	0	9029

Download and subset

Canadian Tobacco and Nicotine Survey, 2021

English ▾

CTNS-5305-E-2021_F1.tab

Statistics Canada, 2023, "Canadian Tobacco and Nicotine Survey, 2021", https://doi.org/10.5072/FK2/ZHT5KX, Borealis, V1, UNF:6:I5CjFBAeehXcIm89HK6dpA== [fileUNF]

< Hide Groups

⊞ Cross Tabulation   ☰ Summary

	ID	Name	Label	Weight
▣				
☑	v6423547	AGEGROUP	Age group of person	
☑	v6423552	ALC_05	Frequency drank alcoholic beverage - past 30 days	
☑	v6423564	ALC_10	Freq drank 4 drinks or more on one occasion - past 12 months	
☑	v6423571	CAN_05AR	Smoked cannabis - lifetime	
☑	v6423570	CAN_05BR	Age first time smoked cannabis	
☐	v6423525	CAN_10AR	Frequency smoked cannabis - past 30 days	
☐	v6423539	CAN_10BR	Nb of days smoked cannabis at least once past month - grouped	
☐	v6423524	CAN_15AR	Frequency smoked cannabis/tobacco mix - past 30 days	
☐	v6423569	CAN_15BR	Nb of days smoked can./toba. mix at least once a week - past 30 days	
☐	v6423529	CAN_17R	Frequency consumed edibles - past 30 days	

All Variables
CAN: Cannabis
DV: Derived Variables
ALC: Alcohol
IU: Initial use
VAP: Vaping
OTP: Other tobacco product status
RD: Reference date
DEM2: Demographics 2
SRV: Survey Related Variables
GDR: Gender
TBC: Tobacco

Search   🔍

Items per page 25 ▾

Download
Download subset
Download original file
Download tab-delimited file
Download RData format file
Download variable metadata

Examples & demonstration

———

Go to: Odesi.ca

"Vaping"

Canadian Tobacco and Nicotine Survey, 2021

"Justin Trudeau"

Canadian Election Study, 2021

Conclusion

———

- Depositing and sharing tabular data in Dataverse can support data exploration
- Using Dataverse APIs you can build data exploration tools
- Data Explorer is an open, ready to use tool
- Data Explorer can be built into integrated data search sites, such as Odesi.ca

For more information about Data Explorer reach out to us at borealisdata.ca or contact me at amber.leahey@utoronto.ca