a place of mind

Finding Survey Data Workshop Exercises

Note*:* These activities work best using **Firefox**. Also, to force Firefox to always ask where to save data you want to download, in Firefox, click on the icon that looks like **three horizontal lines** (top right) and select **Options**. Under **Downloads**, select **Always ask me where to save files** and then close the tab to save the changes.

Exercise 1 – Searching/Browsing for Canadian Survey Data

1. Try searching (or browsing) for surveys by name or subject, using some of the sources below. For example, try searching for specific surveys, such as the "General Social Survey" or "Canadian Community Health Survey", or for topics, such as "Health" or "Education," or try searching for your own area of interest. Compare your results.
   a. **Summon** (http://www.library.ubc.ca/): Search in the **General** tab of the library catalogue to search Summon. Remember, to filter your search results to **Data Set**, by going to the left-hand filters menu; under **Content Type**, click on **More…** and then on **Data Set** and then click on the **Apply** button (near the top of the filter list).
   b. **Abacus** (http://dvn.library.ubc.ca/dvn/dv/UBCLDS): The best way to access Abacus is by going to the library homepage (http://www.library.ubc.ca/). Search for "Abacus" in the **Indexes & Databases** tab. This way, you'll be prompted to log in, giving you full access to download the data you find in Abacus. You can browse Abacus by subject heading on the left, by expanding the tree and selecting a topic of interest, or try searching, using either the search box (top right) or the **Advanced Search** by selecting that option underneath the search box.
   c. **Nesstar** (http://nesstar.library.ubc.ca/webview/): Expand the tree on the left to see our data categories and the surveys they contain. Browsing, not searching, is recommended here.
   d. **<odesi>** (http://search1.odesi.ca/#/): Use the search options to run some searches, or click on the **Browse** button at the top to browse the category tree to find relevant surveys.

2. Try searching for variables to find out which survey will contain that information, using the same sources (except Summon). For example, try searching for "smoking" or "volunteering," or try searching for your own area of interest. Compare your results.
   a. **Abacus**: Use the **Advanced Search** option and try searching next to **Variable Information Contains.**
   b. **Nesstar**: Expand a survey and year of interest, and then click on the **"table"** icon, then expand **Variable Description** to browse the list of variable categories and variables.

Finding Survey Data Workshop Exercises

c. **<odesi>**: Type in a search term, and then from the drop-down next to the search box, select **Variable/Category Label** to search at the variable level. You can also browse variables similarly to Nesstar, by going into the **Browse** view.

Exercise 2 – Downloading Canadian Survey Data

If the survey data you are looking for can be found in Nesstar (ours or Statistics Canada), use Nesstar to download the data, otherwise, use Abacus.

1. Let's download the full dataset for a General Social Survey cycle from Nesstar in SPSS format, along with its documentation. Go to **Nesstar** (http://nesstar.library.ubc.ca/webview/) and browse to **Statistics Canada, General Social Survey (GSS), Cycle 27, 2013: Giving, Volunteering and Participating**, and then click on the **"table"** icon next to **General Social Survey, Cycle 27: Giving, Volunteering and Participating (GSS GVP)**.
2. Click on the **"download"** icon (looks like a floppy disk) from the top right menu.
3. From the **Select Data Format** drop-down menu, select **SPSS**. Make sure that **Include documentation** is checked. Click on **Download**.
4. At this point, you will be prompted to login with your barcode (*found on the* back of your UBCCard or Library card) and PIN. *Note: Nesstar only accepts upper case letters for UBC passwords (PINs). If your PIN has lower case characters, please enter them as upper case.* More information about your PIN, including where to go to change it can be found here: http://services.library.ubc.ca/borrowing-services/library-account/
5. Once logged in, you should be taken back to the download screen where your download should start automatically (there is a link near the top you can click if this is not the case). Your files will be packaged as a zip file. Save it somewhere you can find it, for example: your documents folder or C:\Temp.
6. Go to that folder and extract the files by right clicking on the zip file, selecting **PeaZip** and then **Extract here**.
7. Double click on the **SAV** file to open up the data and view it in SPSS. In SPSS, you should see two tabs near the bottom, so you can switch to the **Variable View** (to learn more about the variables/columns) and to the **Data View** where you will see the actual survey data. The accompanying **PDF** file contains important documentation to help you understand the data.

    Note: You can use these steps to download the data in any format listed in Nesstar, including Stata.

8. These steps allow you to download the entire data set, but in some cases, you may want to only download a subset of the data. If you go back to Nesstar's download screen, instead of clicking on **Download**, click on **Subset.**

9. Expand your variable list on the left to find the variables you want to include in your subset. To include a variable, click on the "graph" icon next to the variable name and select **Add to subset** from the pop-up menu. You should see the variable name appear in the box and a note at the bottom should say **Filter is on**. Add a few variables. When you are done adding variables, click on **OK.**

10. Now when you click on **Download**, you are actually only downloading the variables you selected. You can verify this by viewing the data, as you did in steps 4-7.

Downloading data from Abacus:

1. Let's download the full dataset for a General Social Survey cycle from Abacus in SPSS format, along with its documentation. Go to **Abacus** (hopefully still open from Exercise 1).

2. Search for **GSS Cycle 27** and then click on the **Giving, Volunteering and Participating (GSS GVP)** record's title. The first tab should present you with information about the data set.

3. Click on the **Data & Analysis** tab. Here you should see a list of files you can download, including documentation and data.

4. Select which files you want to download, and either click on **Download** next to each file, or use the **Download Selected Files** button at the top to bulk download your files. For example, for SPSS data, download the syntax file, **GSSC27GVPpumf_syntax_e.sps**, and its corresponding data file, **GSS_ESG_C27_GVP_PUMF_FLAT.zip**. You should also download the documentation, if you want to make sense of the data, starting with the User' Guide: **GSS_ESG_C27_GVP_FMGD_gid_En.pdf** Again, save these files somewhere you can find them, for example: your documents folder or C:\Temp, and extract any zip files.

5. You can then use the syntax/command file (i.e., the SPS file) to prepare your data for SPSS. Start-up **SPSS** (there should be an icon on the desktop). Then select **File -> Open -> Syntax** and chose the .sps file you just downloaded.

6. This file is a series of commands that SPSS will run on the accompanying data file before bringing in that data into SPSS. All we need to do is tell this script where the data is. So we need to edit the path to point to our data file. Go back to where your data file is in Windows Explorer – click on the top location bar to get the path, copy and then paste it into the .sps script replacing the path listed next to **FILE=** (keeping the last part – the data file name: \GVP_PUMF_MAIN.TXT)

7. If you then scroll down to the very bottom you'll see another command that saves an SPSS file – edit this path as well.

Finding Survey Data Workshop Exercises

8. Then select **Run -> All** to run all the commands in the script. If all goes well, your data and variables descriptions should all appear in SPSS. For more details on working with syntax files, see http://www.icpsr.umich.edu/icpsrweb/NACJD/support/faqs/2007/02/how-do-i-use-spss-setup-file-to-import.

Exercise 3 – Online Analysis Using Nesstar

1. Let's use the online analysis tools in Nesstar with the 2010 Canadian Community Health Survey data. Go to **Nesstar** (http://nesstar.library.ubc.ca/webview/) and browse to **Statistics Canada, Canadian Community Health Survey (CCHS), 2010**, and then click on the **"table"** icon next to **Canadian Community Health Survey – Annual Component, 2010**.
2. Expand your variable list on the left to find a variable of interest. If you click on it, you will see more information about the variable, including frequencies and summary statistics. For example, click on **Frequency of having 5 or more drinks** under the variable category, **Alcohol Use**.
3. Next, let's try recoding the age variable from multiple categories down to three: 12-29, 30-59, 60+. Click on the **"compute"** icon (looks like a Sigma – sum operator) from the top right menu. (If prompted to login again, do so.)
4. From the **Please select a type of calculation**, select **Recode**. (You can see from the options that you can create new variables, not just from recoding, but also from other mathematical operators, such as adding multiple variables together, multiplying, etc.)
5. From the **Dwelling and household variable**s category, click on **Age** and select **Recode variable**.
6. If you scroll down, you'll see the existing categories for this variable. If we want to create a new category for ages 12-29, we need to combine the categories ranging from 1 to 5, so for the **Old Value** click on the radio button next to **Range**, and then enter in "1" through "5". For the **New Value**, next to Value, put "1", and next to **Label**, put "12-29". Then click on **Add.**
7. Using these steps as a guide, create categories for 30-59 and 60+.
8. Finally, give your new variable a label, such as "AgeRecoded" and then click on **OK.** You should now see it in the variables list at the bottom under **User defined variables**. If you click on it, you'll see its frequencies and summary statistics, and you can now use this variable in any other online analysis you're interested in. This variable only shows up in your Nesstar account, which is why you are prompted to login when doing online analysis in Nesstar.
9. Next, let's try creating a cross tabulation, showing our alcohol consumption variable by province. First click on the **Tabulation** tab.
10. Next click on the variable **Frequency of having 5 or more drinks** under the variable category, **Alcohol Use**, and select **Add to row**. Then click on the variable **Province of residence of respondent** under the variable category, **Geography variables**, and select **Add to column**. You should now see a cross-tab of alcohol use by province. If you would like to visualize this

information, you can do so by selecting a type of chart from the **"Barchart"** icon from the top right menu.

11. So far, we haven't considered our weight variables in any of our online analysis. If we want to apply a weight variable, we can do so by clicking on the **"weight"** icon (looks like a scale) from the top right menu. Select **Weights – Master** and then the right-pointing arrow key to move it to the **Weighting variables selected** box on the right. Then click on **OK.**

12. Your cross-tab numbers should have changed slightly to more closely represent the population, instead of the sample. At the bottom of the table, you should see the message **Weight is on** to know when a weight variable is applied.

13. Finally, let's look for a correlation between two variables: suicide and depression. Click on the **Analysis** tab and select **Correlation.**

14. Next under the variable category **Suicidal thoughts and attempts**, click on the variable **Has considered suicide - lifetime** and select **Add to correlation**. Then under the variable category **Depression**, click on the variable **Depressed >= 2 weeks – 12 mo** and select **Add to correlation**. You should now see correlation information between the two variables in the table.

This is just some of the basic functionality of Nesstar. For more information, you can read the Nesstar help by clicking on the **"help"** icon (looks like a question mark) from the top right menu. Feel free to also browse the tool to see what other options are available.

Exercise 4 – Searching and Downloading from ICPSR and SDA

1. Let's access ICPSR by going to the library homepage (http://www.library.ubc.ca/). Search for "ICPSR" in the **Indexes & Databases** tab. If you haven't signed up for an account with ICPSR before, do so now by following the instructions in the yellow box under the ICPSR entry. Your account should be created instantaneously. If you already have an account, just click on the ICPSR link.

2. From either the account sign up form or the main page of ICPSR, select **Find & Analyze Data**.

3. Try browsing some of the datasets to get a feel for what is in ICPSR. Try running a few survey level searches here, and also some variable-level searches in the **Search/Compare Variables** section by selecting it from top menu.

4. Feel free to browse the other sections of the website, especially the **Resources for Students** and the **Get Help** sections.

5. Next, let's search for "General Social Survey" in the **Find Data** section. From the first result (which should say General Social Survey, 1972) click on **General Social Survey Series** underneath this entry.

Finding Survey Data Workshop Exercises

6. From the list of all US General Social Surveys available, select **General Social Survey, 1972-2012 [Cumulative File].** This page will provide you with more information about the survey (including an example citation), along with links to download the documentation and data.

7. Next, let's look at the online analysis tool, SDA, and use it to start exploring and downloading a subset of this survey's data. Click on **SDA** under the **Datasets** heading**.** You may be prompted to log into your account at this point and accept the terms of use.

8. You can browse the categories of variables to find a variable of interest. Select a variable to copy its variable name into the top box. Click on **View** next to it to see frequencies and other information about the variable. For example, take a look at **DRINK** under **Societal Concerns, Substance Consumption.**

9. Next, try to download a subset of the data. From the top menu, select **Download** and then **Customized Subset**.

10. Choose your type of data file, whether you want a codebook that only contains information on the subset of variables, and if you want to create/download any syntax/command files for a particular software package. For this example, select **SPSS**. You can also apply filters, if you wish. Click on **Selection Filter(s):** for more information.

11. Then you can either specify the names of the variables you want (if you know them, such as DRINK or AGE) or you can select groups or specific variables from lists. For example, let's select **ALL** next to the **Societal Concerns** group of variables to download all of those ones. Then let's select **SOME** next to the **Personal and Family Information** group. You'll get a chance to say which variables you want from that group on the next screen. Scroll to the bottom and click on **Continue**.

12. While holding down the **Ctrl** button, click on the variables you want. For example, try selecting **MARITAL** and **REGION.** Then click on **Continue**.

13. You'll be presented with a screen summarizing your selections. If you're happy with it, click on **Create the files**. You'll be presented with a screen with links you can use to download the data and documentation. It might be helpful to rename the files to "Data.txt" and "Syntax.sps".

14. Once downloaded, use the steps in the **Download data from Abacus** section of **Exercise 2** to bring the data into SPSS (page 3 of this handout).

Exercise 5 – Online Analysis Using SDA

1. Next, let`s create a crosstab showing DRINK vs REGION. From the top menu, select **Analysis** and then **Run Frequency and Crosstabulation**.

2. Add **DRINK** under **Societal Concerns, Substance Consumption** as a row, by clicking on it, but this time, instead of selecting **View** next to the top box, select **Row** just underneath it**.** Add **REGION** under **Personal and Family Information, Demographic Variables**, similarly, but as the column.

3. Then click on **Run This Table** on the right (might need to scroll down a bit) to see your crosstab.
4. Finally, let`s look for a correlation between two variables: age and health. From the top menu, select **Analysis** and then **Correlation Matrix**.
5. Click on **AGE** under **Respondent Background Variables, Age, Gender, Race, and Ethnicity**. Then click on **Vars to Correlate**, next to **Copy To:**, underneath the variable name copied to the top box. Do the same for **HEALTH** under **Personal Concerns, Health and Happiness**.
6. Then click on **Run Correlations** on the right (again might need to scroll down a bit) to see the correlation information.

Note: For exercise 5, we did not apply any weight variables just to simplify this example; however, for any real-world analyses, we would need to consider weighting and consult the codebook to understand which weight variable to apply.

This is just some of the basic functionality of SDA. For more information, you can click on **Getting Started** at the top, or any of the in-context **General** help link (or other links) throughout the tool for specific help. Feel free to also browse the tool to see what other options are available. If you head back to the main ICPSR page, you will see a link called **Teaching & Learning with ICPSR**, which also has very helpful exercises you can work through to learn more about ICPSR, social research and SDA.

Useful Links

Online Guides

- Data Guide: http://help.library.ubc.ca/finding-resources/finding-data/
- How to Cite Data Guide: http://help.library.ubc.ca/evaluating-and-citing-sources/how-to-cite-data/
- How to Cite Statistics Canada Products: http://www.statcan.gc.ca/pub/12-591-x/12-591-x2009001-eng.htm
- ICPSR – Data Citations: http://www.icpsr.umich.edu/icpsrweb/content/shared/ICPSR/faqs/why-and-how-should-i-cite-data.html

Social Research Resources

- Guppy, Neil and George Gray. (2008). *Successful Surveys: Research Methods and Practice. Fourth Edition.* Toronto: Thomson Nelson. [Book]
- Neuman, W. Lawrence and Karen Robson. (2012). *Basics of Social Research: Qualitative and Quantitative Approaches. Second Canadian Edition.* Toronto: Pearson Canada, Inc. [Book]

Finding Survey Data Workshop Exercises

- Bryman, Alan. (2012). *Social Research Methods. Fourth Edition.* Oxford: Oxford University Press. [Book]
- Teaching & Learning with ICPSR: http://www.icpsr.umich.edu/icpsrweb/instructors/index.jsp

Statistics and Statistical Software Resources

- Research Commons (workshop and consultations on SPSS): http://koerner.library.ubc.ca/services/research-commons/
- Statistical Consulting: http://www.stat.ubc.ca/SCARL/
- There are many online tutorials and resources on statistics and statistical software. Examples:
  - Princeton University Library Guide: http://libguides.princeton.edu/c.php?g=84085&p=543671
  - MIT guide: http://libguides.mit.edu/c.php?g=176427&p=1159958

Data Sources

- Summon: http://www.library.ubc.ca/ (Note: Filter by data set)
- Abacus: http://dvn.library.ubc.ca/dvn/dv/UBCLDS) (Note: Go to library homepage (http://www.library.ubc.ca/). Search for "Abacus" in the Indexes & Databases tab.)
- Nesstar: http://nesstar.library.ubc.ca/webview/
- <odesi>: http://search1.odesi.ca/#/
- Statistics Canada's Nesstar: http://dli-idd-nesstar.statcan.gc.ca/webview/
- Statistics Canada's Survey Data Search: http://www120.statcan.gc.ca/dli/e1/stu
- ICPSR: http://www.icpsr.umich.edu/icpsrweb/landing.jsp (Note: Go to the library homepage (http://www.library.ubc.ca/). Search for "ICPSR" in the Indexes & Databases tab.)

Feel free to contact me for any questions about finding, accessing or citing data!

Kelly Schultz, Data & GIS Librarian
Koerner Library
kelly.schultz@ubc.ca
604-822-8667