

Analysing Census At School Data

www.censusatschool.ie

Project Maths Development Team 2015

Excel for Analysing Census At School Data

Having retrieved a class's data using CensusAtSchool:

1. Save the downloaded class data which is a .csv file as an .xls file.
2. Delete any unwanted variables such as the teacher name, date stamp etc.
3. Freeze the top row so that one can scroll down to view all of the data while the top row remains in view.
4. Select certain columns (not necessarily contiguous) for use as a class worksheet. Aim to have a selection of categorical data (nominal and ordered) and numerical data (discrete and continuous) unless one requires one type exclusively. The number of columns will depend on the class as too much data might be intimidating for less experienced classes. Students will need to have the questionnaire to hand to identify the question which produced the data.
5. Make a new worksheet within the workbook of downloaded data using the selected columns.
6. Name the worksheet i.e. "data for year 2 2014" or "data for margin of error investigation".
7. Format the data to show all borders (useful if one is going to make a selection of the data and put it into Word).
8. Order selected columns of this data perhaps (and use expand selection) so that rogue data such as 0 for height etc. can be eliminated. It is perhaps useful for students to see this. (Use Sort & Filter.)
9. Use a filter on data, for example based on gender, to generate further worksheets for future use
10. Print the worksheet.
(Alternatively select all the required data and insert into Word (Page layout: Landscape; Margins: Narrow))
11. Repeat steps 3 to 9 to give two different sets of data which when printed back to back give data for use in class and data for use as homework.
12. Using the random data selector, data from questions which are common from year to year, could be used for comparison type questions. Likewise two teachers with different class groups could swap their data for use for comparative purposes.

It might also be useful to show how to:

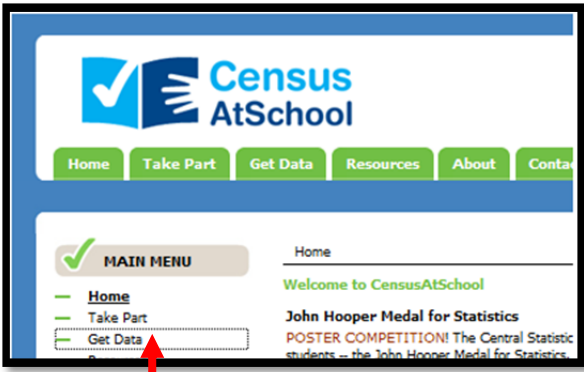
- Change the width of a column
- Hide a column
- Convert data to number format
- Use the "COUNTIF" command in Excel to facilitate making a table if there is a lot of data
- Draw a chart in Excel (in particular pie-charts and comparative bar charts, which is not possible in GeoGebra).

Data Tool Tasks



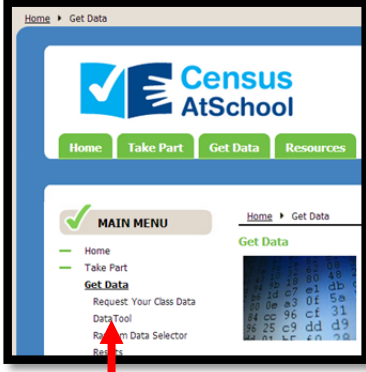
The **DataTool** allows you to choose a random sample of data from the *CensusAtSchool* database, to select variables and then to analyse this data. It can provide both graphical and numerical summaries of data. It can show sampling variability. It allows you to make comparisons e.g. compare girls to boys etc., select sample features and investigate relationships between variables. The DataTool works through a web browser and requires a live internet connection.

Accessing the DataTool: Open www.censusatschool.ie

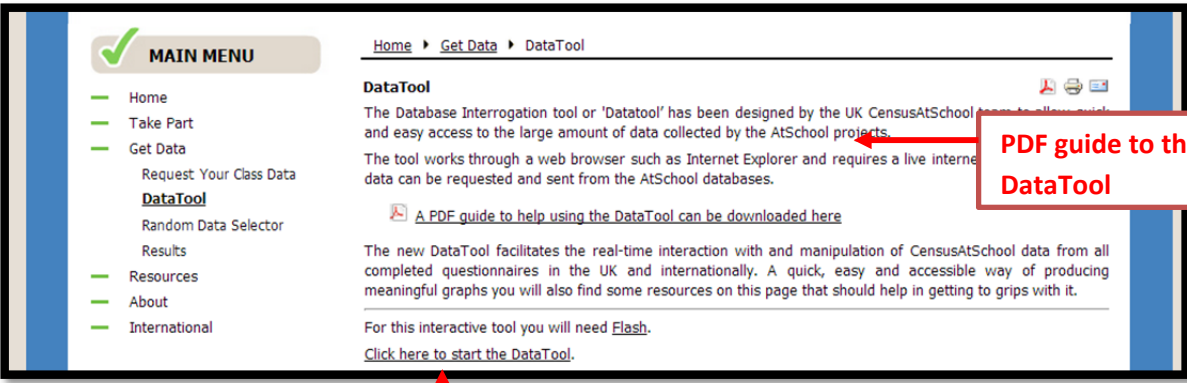


Cli

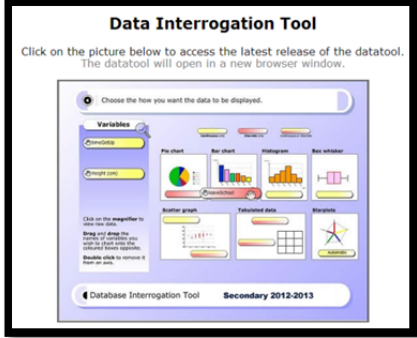
Select "Get Data"




Select "DataTool"



Select "Click here to start the DataTool."



Click on the picture to access the DataTool.



Click on the logo to start.

Data Tool Tasks

Click on the flag of the country whose data you wish to analyse.



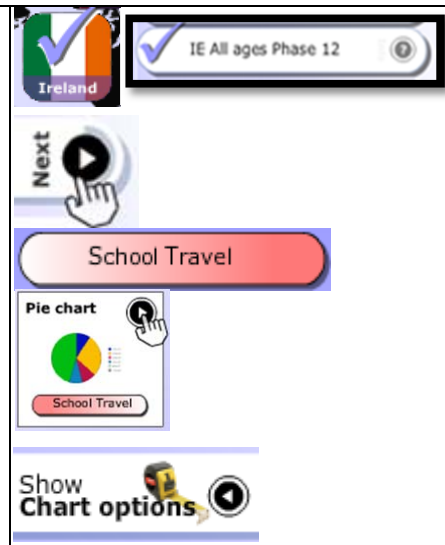
By clicking on the Irish flag, the data from completed Irish questionnaires becomes available.



Task 1 involves analysis of categorical data.

Task 1: How do Irish secondary students travel to school?

- i. Choose data from Ireland, **IE All Ages Phase 12**
- ii. Click **Next**
- iii. Pick up the **School Travel** tab and put it on the left hand side (under **Variables**) and click **Next**
- iv. We want a random sample of size 100 so click **Next** again
- v. Now pick up the **School Travel** tab and put it under the Pie chart. Click the arrow next to the pie chart and a pie chart displaying the data should appear.
- vi. Click on "Show Options".
- vii. Click on Chart Options to have the results arranged in order.



Q 1: Based on this sample, what is the most common method of travel to school by these students?

Q 2: Make a table showing the proportion of students who travel to school in each category. Do these figures add up to 100%? Explain your answer.

Q 3: Can you identify a source of bias in the sample results?

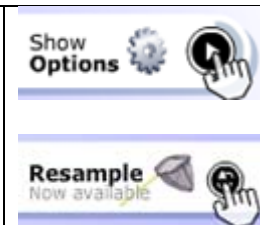
Q 4: Compare the findings for this sample with the findings for the population of 8,610 students from 243 schools in Ireland who completed the 2012/2013 CensusAtSchool questionnaire, as shown below:

- The most common method of travelling to school by students was by Car (47%). Travelling by Bus (27%) and Walking (21%) were the next most popular methods of travelling to school.

<http://www.censusatschool.ie/en/get-data/results/phase-12-1213>

Q 5: Use the **Resample** button to see if another random sample of 100 students gives a similar answer to the first sample? (This is in **Show Options, Resample.**)

Try this number of times. (You have to wait 20s between samples). What do you notice?



Q 6: Use another graphical representation to summarise how a sample of 100 students travel to school. Use the Back button, double click on the Travel to School tab and move it to a different graphical representation. Which graphical representations do you think is best? Justify your answer.

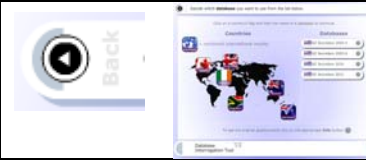
Q 7: Compare this data with your own class data. What graphical and numerical representations can you use?

Q 8: Compare this sample data with a sample of data on methods of travel to school from New Zealand or the UK.

Data Tool Tasks

Task 2: Analysis of categorical data (comparative descriptive)

Click **Back** until you get to this webpage.



Task 2: Do boys like the same colour as girls?

- Choose **UK Secondary 2004 –5** data
- Click **Next**
- Choose the **Favourite colour** tab and put it on the left hand side under **Variables**
- Click **Next**.
- As we want to compare boys' and girls' answers we need to choose **Part of database** and **Two samples** (Click on Part of the Database and One sample if you only want to look at the data from Boys for example or from students of a particular age etc.)
- Click **Next**.
- We want **Sample A** to be girls so click on **Gender** and on the left hand side choose **Females**.
- Click **Next**
- Click on **Gender** and on the left hand side choose **Sample B** to be **Males**.
- On the graph page select either Bar charts or Pie charts.
- Which is the best graphical representation of the data? Explain.
- Resample a few times.

Question: What have you found out?


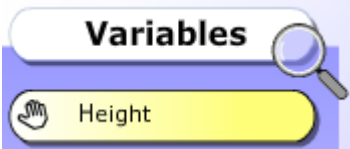
Task 3: Analysis of numeric continuous data (summary descriptive)

Task 3: Describe the distribution of a sample of Irish students' heights.

Use graphical and numerical summaries of the data.

- Click back to Countries page.
- Choose **IE All ages Phase 11** data. Click **Next**
- Choose the **Height** tab and put it on the left hand side under **Variables**. Click **Next**.
- Choose a sample size of **250, Part of database, One sample**
- Click **Next**. You now can select the features you want the sample to have.
- Click on height (which has an asterisk beside it). Set sensible values for the maximum and minimum heights as there may be rogue data values. Click on **"Save this option"**.
- Click on **Gender** and select **Females**. (Other filters may be applied also e.g. age or age range, only those who eat breakfast etc.). Click **Next**
- Choose a graphical representation for the data. Describe the distribution.
- Select **Show Options** and **Chart Options** to change **the scale** and **number of bins** for a histogram. The y-scale may need to be adjusted if the histogram is too high. **Resample**.
- Use the **Back button** and display the data using a different graphical representation. Investigate the **Chart Options** and **Resample**.

Data Tool Tasks

<ul style="list-style-type: none">From the page with the graph you chose, click the Back button to go back to the graphs page	
<ul style="list-style-type: none">Click on the magnifying glass beside the word Variables, to view the original data, a summary table and the summary statistics. (You can copy this data into GeoGebra or Excel if you wish.)	
<ul style="list-style-type: none">Investigate how sample size affects the variability of sample statistics.	

Task 4: Investigate the relationship between variables (Bivariate analysis)

Task 4: Investigating relationships between variables

- Click back to Countries page.
- Choose **IE All ages Phase 9** data

Investigate :

Is there a relationship between:

- Height and foot size?
- Height and length of ring finger?
- Height and best time on the interactive quiz?
- Height and whether or not you think the 2012 Olympics is good for your area of the country?

Investigate using a suitable graphical representation. Investigate the summary statistics.

Is the relationship positive or negative, strong or weak?

Data Tool Tasks

Using your own data:

To use your own data :

- Select the options shown.
- (If the data has not been “cleaned” use “Part of database”)
- Copy the data from an Excel spreadsheet.
- Select “Next”.
- Paste the data into “Extra user data” and proceed to data analysis as in previous tasks

The screenshot shows a software interface for data collection. At the top, it says "You can select the type and size of random sample you wish to collect." Below this, there are two main sections: "Variables" and "Random sample".

Variables: A list of variables is shown, with "Media Story" selected and highlighted in red.

Random sample: A section for selecting the size of the random sample (from 50 to 250) with a dropdown menu set to "0". Below this, there are four options for sampling:

- Whole database:** Collect sample(s) of your variables from anywhere in the database. (Checked)
- Part of database:** Collect data selected for particular groups or categories such as boys or girls only.
- One sample:** Collect just one sample (A) from the database to investigate. (Checked)
- Two samples:** Collect two samples, A & B, to investigate and compare.

At the bottom of the "Random sample" section, there are two buttons: "Use my own data" (checked) and "Use my own data as sample B".

Below the "Random sample" section, there is a "Next" button with a play icon and a hand cursor. At the very bottom, there is a large orange button labeled "Extra user data".

Learning outcomes:

- Awareness of how different data types determine the graphical representations used
- Experiencing sampling variability and how it is affected by sample size (Syllabus 1.7 – Recognise how sampling variability influences the use of sample information to make statements about the population.)
- Seeing and analysing the effect of using different sized intervals (all of equal size) when plotting histograms
- Describing the shapes of different data distributions
- Awareness of the different types of questions
- Investigating descriptive type questions, comparative type questions and relationship type questions

Census at school Statistics Resource 1: Getting to school

<http://www.censusatschool.ie/en/resources/strand1statistics/154-getting-to-school>

See also: www.cso.ie

<http://www.cso.ie/en/media/csoie/census/documents/thisisirelandpart2census2011/This,is,Ireland,Highlights,,P2,Commentary.pdf> (Page 34)