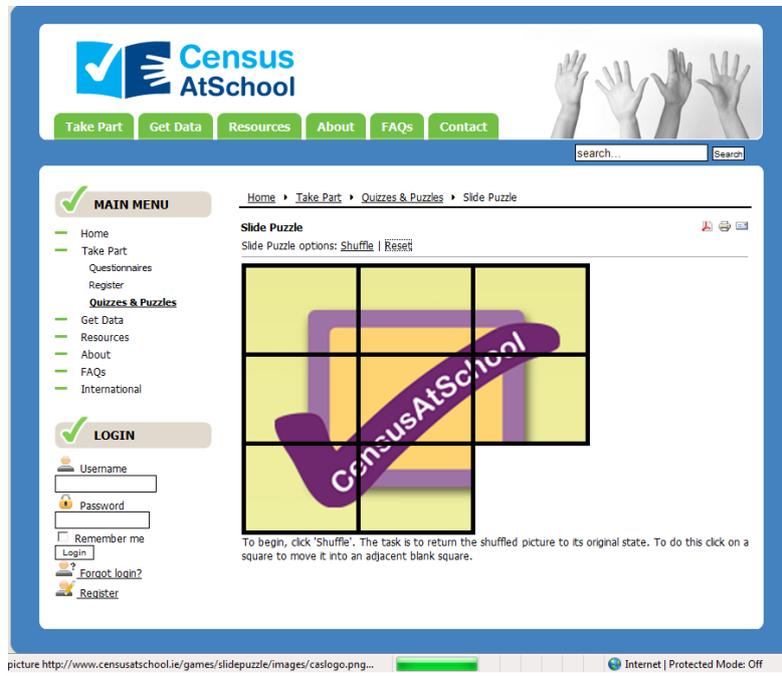


## Statistics and Data Handling – a real data approach using CensusAtSchool



The screenshot shows the CensusAtSchool website interface. At the top, there is a navigation bar with the following links: [Take Part](#), [Get Data](#), [Resources](#), [About](#), [FAQs](#), and [Contact](#). A search bar is located to the right of these links. Below the navigation bar, the main content area is divided into two columns. The left column contains a **MAIN MENU** with links to Home, Take Part, Questionnaires, Register, Quizzes & Puzzles, Get Data, Resources, About, FAQs, and International. Below the menu is a **LOGIN** section with fields for Username and Password, and buttons for Login, Remember me, Forgot login?, and Register. The right column displays the **Slide Puzzle** game. The game interface features a 3x3 grid of squares. The top row contains a yellow square, a purple square, and a yellow square. The middle row contains a yellow square, a purple square, and a yellow square. The bottom row contains a yellow square, a purple square, and a yellow square. A large purple watermark with the text "CensusAtSchool" is overlaid on the grid. Below the grid, there is a text box that reads: "To begin, click 'Shuffle'. The task is to return the shuffled picture to its original state. To do this click on a square to move it into an adjacent blank square." The browser's address bar at the bottom shows the URL: <http://www.censusatschool.ie/games/slidepuzzle/images/caslogo.png...>

## Statistics and Data Handling – a real data approach using CensusAtSchool

We live in an information society where decisions in politics, business, marketing, education, science, transportation, and health care are made on the basis of data which has been collected, analysed, interpreted and acted upon. Current data is used to predict how many schools, and hospitals we will need in 10 years time and also what jobs we need to train people for, now, so that they we be ready to take advantage of changes in the economic climate and supply the needs of a changing society.

Basic statistical skills – reading and creating tables, constructing graphs and calculating measures of central tendency (mean, median and mode) – are required across the curricula. Students practice these skills not only in the traditionally number-based subjects like mathematics and economics, but also in history, geography, business, science, law, health and safety, research methods and other subjects. Data management skills and data managers are crucial in our information age. There is an increasing amount of employment in this area,(thousands of jobs on jobs.ie require data analysis skills) not to mention a need for citizens to able to critically evaluate the statistics they will encounter on an everyday basis.

When students use data about themselves, or their class, they are drawn more deeply into thinking critically about the information. They find the exercises real, interesting and enjoyable. They compare their findings against their own experience. They question, draw conclusions and ultimately construct their own learning. When students are analyzing and interpreting data, they are engaged in thinking critically about what they are studying.

In any statistical investigation students need **a question to be answered** and **data to help answer the question**.

**CensusAtSchool** provides both – based on the questionnaire, students can come up with a variety of questions to be answered. They have their own class results from the questionnaire as a rich source of data, and they can compare their findings with data from students of their own age in different countries – connecting people through maths! As the questions provide different data types both categorical (nominal and ordered) and numeric (discrete and continuous) students have examples of the different types which they can directly relate to.

Having produced their own data, students can proceed to organise and describe the data (descriptive statistics) and draw conclusions from the data (statistical inference).

## Census at School

### A real data approach to teaching statistics

Students complete a brief online survey, analyze their class results and compare themselves with students from other countries e.g. Canada, Australia, South Africa, New Zealand and the UK. The project began in the UK in 2001 as a way of getting young people and the population, aware of the Census and its importance to our lives.

The students fill in a survey online anonymously, answering non - confidential questions. Their responses become part of a national database, which in turn becomes part of an international database (with over one million entries) which is maintained in the UK and this year Irish data will be processed in Ireland. Census at School is a voluntary international project which will not be used for research purposes.

The class data is returned to the teacher, on request from the headquarters of censusatschool in the UK in spreadsheet form, from which students can pick a dataset to enable them to carry out some statistical investigation. They can use the data to answer a question – e.g. is there a greater variation in height for 15 yr old boys as opposed to 15 year old girls, is there a correlation between height and arm span (vitruvian man), what is the real level of interest in reality TV out there among 16 year olds etc. Datasets from the international database may also be requested for comparative purposes using the “**Random Data Selector**”. Do the findings of my class agree with other those of students in other countries?

#### To Register with censusatschool and take the questionnaire

All instructions can be found on the project maths website in a leaflet using the link below.

[http://www.projectmaths.ie/resources/census\\_at\\_school\\_leaflet.pdf](http://www.projectmaths.ie/resources/census_at_school_leaflet.pdf)

Go to [www.censusatschool.ie](http://www.censusatschool.ie)

On the homepage click on the Take Part link on the main menu, then “Register”. Click on “Registration form” which you print out and complete and post back to NCTE. You will receive back an email with your personal username e.g. Murphy 4522. Students will need your personal username and the school roll number when they are filling out the questionnaire online. Under “Take Part” on the main menu there is also a link to Questionnaires and if you click on Phase 9 (latest and with some questions with an Irish theme) you can either download the questionnaire or your students can take the questionnaire. Clicking on the Get Data link on the main menu, there is a link here which leads to access to the **Random Data Selector**. You only need an email address to access this data. This allows you to access data from students in other countries which is sent to you in spreadsheet form and students can use it to make comparisons with their own data. The Get Data link will also allow you to retrieve your data in spreadsheet form when your students have filled in a questionnaire online.



The project is on Phase 9 (Phase 1 was the first yr of the project) at the moment (started in the UK in 2001 before the Census was collected) but all previous phases are still alive and accessible from the classroom materials section of the website, should you decide that you prefer any of those questionnaires. Each questionnaire has some questions in common and then the rest of the questions may vary from interest in sport, to the environment, to nutrition, music tastes and this year there is a focus on the upcoming Olympics. It might be an idea to give a sample of the questionnaires to students and ask a class committee or student council representatives to decide on which phase to choose.

When the data returns to you in spreadsheet form.....

**TO PROTECT DATA:**

When the dataset is returned PROTECT IT so that pupils cannot change the original data by mistake or design

In EXCEL, give the data Read Only protection with a password that only the teachers know.

When your school's data is in one file, choose *File/SaveAs*, click *Tools, Options* and enter a password in the "Password to modify box". (The password must be exact case and can be composed of letters, numbers and symbols). Click *OK* and you will be asked to re-enter the password and then click *OK* and *Save*.

You need to confirm the replacement of the original file by clicking *YES* and **to activate the protection the file must be closed and then reopened.**

## A broad overview of how the data might be used for statistical investigations

### CLASS 1 Statistics in real life and producing data

Ask students to give instances of where statistics are used to help them or the government or people in general to make decisions. Introduce the idea of sampling and how important it is that it is random.

- Why buy Benecol – claim is that it is proven to reduce cholesterol ( in one ad. 153 people were tested in the sample)
- Why buy Whiskas- 4 out of 5 cats prefer Whiskas !( how many cats in the sample)
- Where will the government decide to build new roads/ schools? Decision will be made on data collected in surveys or from the census etc.
- Why is Coronation Street still being broadcast after so many years? According to figures published in papers 16.8 million people watched Coronation Street on one particular night – a lot of people! How do they know? Was someone looking in through your window that night? Did anyone ask you? Is there some secret device on your TV for recording your TV watching habits? The method is “sampling”, and its proximity to the truth depends on how large the sample is, whether or not its random, if everyone tells the truth, what else was happening that night, and if this was a typical episode etc.!
- Read the small print and beware of the possible abuse of statistics. Remember – “lies, damned lies and statistics” (Disraeli) and “Statistics never lie but liars use statistics”

### TELL PUPILS ABOUT CENSUSATSCHOOL AND THE QUESTIONNAIRE

Inform students about the types of questions involved in the questionnaire. Emphasise that they will be anonymous when filling out their data. They should be aware that they will be able to compare their data with other countries ( using the random data selector mentioned earlier).

Give them a hard copy of the questionnaire. It is recommended that students carry out the measurements before filling out the questionnaire online. It is possible to print out a Word version of the questionnaire and have pupils fill in most of the questions this way so that they will be more efficient when online. Some of the questions are interactive so they need to be online, but with good preparation it shouldn't take more than 15 minutes per student to fill out online. Some students could fill out the questionnaire at home.

Q4 deals with measuring – height, foot length and arm span. **Show them exactly how to do the measurements**, and if there is not enough time in school, depending on the skills base of the student to do this unsupervised, they could do them at home, and have them ready to fill into the questionnaire.

Ask students what information they might find out from the data in the questionnaire, and what they themselves would like to find out from it. Ask them to write down questions they want to investigate when they get the data back, and to predict answers to their questions. They can come up with hypotheses that they wish to test.

When they have filled the questionnaire out online later they may wish to modify these questions.

## INTRODUCE PUPILS TO THE DATA HANDLING CYCLE

Data handling cycle:

Pose a question and predict an answer, collect data, analyse the data, interpret the results and back to look at the question again to see if it has been satisfactorily answered.

**(Issues which will arise along the way:** What data is needed to answer my question, what type of graphical analysis will best reveal patterns in the data, are there other possible interpretations of the results? Do the results show correlation for bivariate data – can this be interpreted as causality?

(If your stress levels are lower when you do yoga, maybe it's because you only have time to do yoga when you are less stressed and not that yoga makes you less stressed.)

## CLASS 2

### **Fill out the questionnaire.**

**Using the data projector, get one or two students (not necessarily the “best” students in the class) to demonstrate how to access the questionnaire, while you tell them what to do, and then they can act as helpers if others have difficulty.**

Students fill in the questionnaire and submit. (Questionnaires include some interactive games such as memory test games or reaction timer testers.) Students have an opportunity to review their responses before finally submitting to the ever expanding database. (Well over 1 million entries).

The teacher explains to the class again what will now happen to their data i.e. will be processed and returned as an EXCEL file.

**Having filled out the questionnaire the students may wish to review or rewrite the questions they wish to have answered when they get the data back.**

**HOMEWORK! THINK ABOUT YOUR QUESTION, AND WHAT DATA YOU WILL NEED. PREDICT AN ANSWER TO YOUR QUESTION.**

## CLASS 3

### VIEWING DATA:

The entire data sheet could be printed out, stuck together and put onto the classroom wall so that pupils can appreciate its size and could identify themselves on it. (Smaller version copied for each student perhaps). However teachers could take certain columns to include all the different types of data and print these on an A4 sheet for each student.

Students will notice codes for certain answers and will need to know how to decode it.

### DECODING DATA

When the dataset is returned DECODE it (**coding sheets supplied by census at school**). If left in coded form students could find data confusing and meaningless. Each question can be called out again by the teacher and the students can view the answers.

### STUDENTS WORK IN PAIRS CHECKING FOR ROGUE DATA:

They could discover the effect of rogue data e.g. if a pupil mistakenly entered their height as 15cm instead of 150 cm, how would this affect the Mean height in say a class of 24 students? They could check heights by first of all deciding what would be sensible limits on heights. Students in pairs could be given responsibility for checking different questions and reporting back.

### PUT THE DATA INTO DIFFERENT CATEGORIES

The teacher informs the students of the different types of data (category (nominal and ordered), numeric (discrete and continuous). Students working in groups then categorise the data and report back to the class where they justify their decisions. This is important as it will determine the graphical representations used (bar charts, pie charts and line plots for category type data and histograms, stem and leaf plots for numeric data and line plots only for numeric discrete. If numeric data is discrete and hasn't a wide range of values a bar chart could be used to describe it graphically but if it has a wide range of values a bar for each value would be cumbersome and a histogram would be more convenient. However histograms are not used for category type data.

### LOOK AT THE RANGE OF NUMERIC DATA SAY FOR HEIGHT.

Range = Maximum - minimum

### BACK TO QUESTIONS THEY HOPED DATA WOULD ANSWER!

#### PREDICT, PREDICT, PREDICT!!

Whatever the information they wish to find from the data, e.g. are girls more interested than boys in reality TV, or are more boys than girls left-handed etc, always ensure that students have predicted an outcome before analyzing the actual data. This generates interest in seeing if their guesses were correct when they do the analysis. When they have come up with their own questions to investigate from the data they are more interested in the outcome.

- Are there many counties represented in the class? ( Bar chart, pie chart or line plot and conclusion using data from Q1)
- What is the distribution of males vs females in the class?(Bar chart or pie chart or line plot +conclusion)
- Test the hypothesis – more children are born in the Summer than the winter using the class as a sample. Use the data on dates of birth and draw a conclusion. Is the graphical representation symmetric or skewed?

Suggestion: Using all category type data each pair of students should pose a question, draw more than 1 type of graphical representation of the data, draw a conclusion and check if data answered the question adequately. State which graphical representation best illustrated the distribution of the data. What was the mode?

#### **CLASS 4 and CLASS 5 Use Numeric data to test hypotheses**

STUDENTS PROCEED TO TEST OUT THEIR HYPOTHESES TO SEE IF THEY ARE TRUE using numeric data.

Histograms and stem and leaf and line plots may be used here but the graphical representation is only one part of the cycle of data handling – students must pose a question to be answered and predict an answer, then having collected the data they analyse it using one of the graphical representations and then interpret the result in light of the question.

The mode, median, mean, range, maximum and minimum can be calculated.

**Histograms** can be drawn for any of the measurements such as height, foot length, length of index finger, length of ring finger.

<http://www.shodor.org/interactivate/discussions/HistogramsVsBarGraph/>

Students will need to decide on appropriate class intervals for the data. Different groups should use different class intervals and compare the graphs obtained. Are the conclusions the same using different class intervals?

<http://www.censusatschool.ie/resources/misc/250-an-interactive-histogram>

<http://www.shodor.org/interactivate/discussions/ClassInterval/>

Students can calculate the mean using class intervals and compare it with the mean using the raw data and explain the difference.

Different groups of students could try different vertical scales on their histograms and compare the information conveyed in each case

<http://www.shodor.org/interactivate/discussions/VerticalScaleDiscuss/>

Students should note if the distribution is roughly symmetric or if it is skewed positively or negatively. Are there **outliers**?

It is very important that they look at the **variability** of the data. The **range** is one measure and then they can be introduced to the standard deviation. **Standard deviation** is another measure of variability. Heights of boys and girls could be treated separately and the standard deviation compared for each distribution and conclusions reached. Students should note outliers and calculate the mean and standard deviation leaving out the outliers. This way they will see that the mean and standard deviation can be greatly affected by the presence of outliers.

Student then find the median, lower and upper quartiles and the interquartile range.

Is the median affected by the presence of outliers?

## CLASS 6

e.g. are the boys in the class taller than the girls – **stem and leaf** would suit this.

A line plot is a nice introduction to a stem and leaf.

[http://www.harlan.k12.ia.us/mrsc/line\\_plots.htm](http://www.harlan.k12.ia.us/mrsc/line_plots.htm)

Depending on their technical skills, the teacher might use this stage to e.g. show how to do a stem and leaf plot. The key issue is that they are working with real data, they have posed a question, predicted an answer and are going to use the data to predict an answer. If the students have the technical skills and are familiar with the various graphical representations, they can be left to decide for themselves, which is the best way to analyse the data. Ultimately **all** students would be given an investigation where they would have to make this decision.

## CLASS 7 and 8 UNIVARIATE VS BIVARIATE DATA

Is there a relationship between height and foot size (scatter plot, noting outliers and finding out if they are due to inaccuracies in measurements or are genuine), This could be done for the whole class or divided by gender. Investigate the strength and direction of the relationship. Calculate the correlation coefficient (HL LC only).

Other possibilities : Age vs height, Age vs foot length, age vs hours spent on social networking sites ( probably won't have enough range of data in one class for this), height vs reaction time, length of index finger vs phone credit

### **Individual projects for pairs of students – come up with 4 questions you wish to answer using the data from censusatschool.**

e.g. the 3 most popular subjects in school, (and then maybe compare it to a similar group of students from South Africa of the same age), (Bar or pie)

Which is the most popular sport in the class? ( Bar, pie or line plot)

Which is the most popular month to be born in in this class? Is it the same for other classes? Is it the same for a random sample of students of the same age from the Southern hemisphere? (bar or pie)

Is there a correlation between length of ring finger and foot size? How strong and in what direction? (scatter plot)

Compare ring finger length for boys and girls (stem and leaf)

What is the distribution of foot lengths in the class? (histogram) etc

When students know the question they want answered, they should identify the data they will need and cut and paste it from the entire data set into a new spreadsheet.

They are required to produce one pie chart, one bar chart, one histogram and one scatter graph of bivariate data. If available they could try this out in Excel or Autograph , and also draw the graphical representations themselves. Students should categorise the data first and then base their graphical representation on the categorisation. If it is numeric data they should supply the following summary statistics mode, median, mean,

ICT – USING EXCEL – useful skill but not a syllabus requirement

Using ICT allows students to focus on **interpreting** data and graphs effectively; because sometimes, the time spent producing graphs can cause them to forget why they are producing them!

ICT methods being learned here are being learned on a “need to” basis as opposed to just as a skill which they may or may not have to use sometime.

Students will need to be able to cut and paste data into new worksheets, use the Graph Wizard, put graphs onto separate pages, paste into a word document, and write a brief report on trends or anomalies noted, plus their reasons, conclusions etc.

They should be able to produce a pie chart, a bar chart and a scatter graph, use MODE, MEDIAN, MEAN (AVERAGE), COUNT, COUNTIF, AUTOSUM functions.

## CLASS 9

Feedback to the class from all the different groups with a general discussion on census at school. How did the class find working with their own data as opposed to data from textbooks? Have you any further questions you would like answered?

## PUBLISH INTERESTING RESULTS IN THE SCHOOL NEWSLETTER OR LOCAL PAPER.

Students can also showcase the results on school notice boards.

Feedback sessions in the classroom where pupils share the outcomes of their investigations and give reasons for their conclusions can throw up some interesting ideas.

An interesting graphic display of an investigation in a local publication could generate good publicity for maths and give it a higher profile in the school. Outings with Maths themes also/ invite in a guest speaker who relies on data analysis to make decisions in their job.

Check out the **Resources** link on <http://www.censusatschool.ie/resources>

Check out [www.censusatschool.com](http://www.censusatschool.com) and you will see that Ireland, Japan and the US are all newcomers to the site

WEBSITES OF UK <http://www.censusatschool.org.uk/>

, CANADA, <http://www19.statcan.ca/r000-eng.htm>

AUSTRALIA, <http://www.abs.gov.au/websitedbs/cashome.nsf/Home/Home>

NEW ZEALAND <http://www.censusatschool.org.nz/>