Data Gems: a guide to dataset sources from Aotearoa and beyond (2023)

Compiled and curated by the *New Zealand Statistical Association's Education Committee*'s *Amy Renelle*, September 2023.

This resource compiles a list of datasets for use by New Zealand teachers and students, fostering statistical and mathematical thinking. Just as the media coverage conceals layers of statistics and maths, this compilation connects you to key local and international sources. These include the *New Zealand Statistical Association*, the *International Statistical Institute*, and global hubs. Dive into a plethora of datasets, complemented by vivid graphics and encompassing health, economic, and social data.

Presented by the New Zealand Statistical Association's Education Committee, this document is intended to empower teachers and students to unearth statistical insights within contexts that resonate with them for classroom use.

Contents

Disclaimer
The descriptions of some of the datasets listed in this guide are summaries derived from the respective websites and do not necessarily reflect the views or opinions of the New Zealand Statistical Association's Education Committee. While we strive to curate and present accurate and up-to-date information, we cannot guarantee the validity, reliability, or completeness of the data presented. Users are encouraged to exercise discretion and to refer to the original sources for comprehensive details.

Datasets from Aotearoa New Zealand sources

Statistics New Zealand
https://www.stats.govt.nz/

Statistics New Zealand is the country's national statistical office, responsible for collecting and providing various statistical information and data about New Zealand's population, economy, society, and environment. The site enables users to download time series datasets from a vast array of topics, build tables from datasets, and find background information about methods such as sampling and inference.
To protect respondent confidentiality, unit-record datasets are not released. However, a large number of tables from these datasets are available and can be explored. Some tables, like information on population, income, etc, by statistical areas, can be explored as if they were unit-record datasets in the statistical areas. The time series datasets are ready for exploration as is.

Keywords: New Zealand data, population statistics, economic indicators, social data, environmental data

data.govt.nz
https://data.govt.nz/

data.govt.nz is a New Zealand government initiative that provides access to a wide range of datasets from various government agencies. It serves as a central hub for open government data, offering datasets related to multiple topics, sectors, and domains. Once you have navigated to the link provided here, scroll down to the list of *Datasets by group* and select an appropriate topic. Many of these are very government-oriented (such as ministry spending data) but some will be of interest to students.
This site is part of NZ's Open Government Data Programme (Open Data NZ). See:
https://www.digital.govt.nz/standards-and-guidance/data/open-data/.

Keywords: New Zealand government data, open data, government agencies, datasets, open government initiative

Figure.NZ
https://figure.nz/

Figure.NZ is a platform that provides access to a diverse range of data and statistics about New Zealand. It aims to make data more accessible and understandable through visualisations and interactive tools, allowing users to explore and engage with various datasets related to New Zealand's economy, society, and environment.

Keywords: New Zealand data, data visualisation, statistics, interactive tools, economy, society, environment

Science New Zealand
http://www.sciencenewzealand.org/

Science New Zealand is a collection of New Zealand's Crown Research Institutes (CRIs) dedicated to scientific research and innovation across various domains, such as agriculture, environment, health, technology, and more. The data available encompasses research findings, scientific reports, datasets, and publications on NZ's scientific progress.

Keywords: New Zealand science research, Crown Research Institutes, innovation, scientific collaboration

Insights by the NZ Herald
https://insights.nzherald.co.nz/

Insights NZ Herald is a dedicated platform that offers in-depth data-driven stories, interactive visualizations, and multimedia content focusing on a wide range of topics, including news, societal trends, economic indicators, and more. Insights NZ Herald provides a thorough understanding of various issues and phenomena around the world.

Keywords: data-driven stories, interactive visualisations, data journalism, NZ insights, data storytelling

Flattening the Curve: Links to statistical or modelling content about CoViD-19
https://new.censusatschool.org.nz/resource/nzsa_edcom_covid-19/

The page provides links to COVID-19 related sites. These resources aim to engage students in understanding and analysing data related to the COVID-19 pandemic, using statistical concepts and methods. The page is on CensusAtSchool and comes from the New Zealand Statistical Association (NZSA) Education Committee.

Keywords: CensusAtSchool, New Zealand, COVID-19, teaching resources, statistical concepts, data analysis

University of Otago - Statistics in Research
https://www.stats.otago.ac.nz/research/Statistics-in-Research/

The University of Otago's *Statistics in Research* webpage offers freely accessible video discussions and accompanying lessons that explore data analysis procedures across a wide range of research applications. The site contains 20 resources. Each has an NZ context communicated by video and text, and the dataset or sets that are part of it. See the introduction to this on CensusAtSchool at:
Datasets-with-their-NZ-contexts_Statistics-in-Research.pdf (censusatschool.org.nz).

Keywords: statistics in research, statistical methodologies, research disciplines, statistical consulting, collaborations, research resources

Charities Services (New Zealand)
https://www.charities.govt.nz/charities-in-new-zealand/the-charities-register/open-data/

Charities Services, a New Zealand government agency, provides open data from the Charities Register. While also available through the data.govt.nz site (above), this open data initiative offers access to a range of datasets related to registered charities in New Zealand. The datasets encompass information about charitable organisations, their activities, financials, and more.

Keywords: New Zealand charities, open data, charitable organisations, activities, financial data, datasets

FluTracking
https://www.flutracking.net

FluTracking is an online health surveillance system that collects and monitors data about influenza-like illness (ILI) symptoms from participants across several countries, including New Zealand. Participants voluntarily report their symptoms, providing valuable data for monitoring and understanding the spread of flu-like illnesses in the community. Students and teachers can contribute by signing up as participants.

ESR Notifiable Diseases Surveillance

https://www.esr.cri.nz/our-research/nga-kete/infectious-disease-intelligence/notifiable-diseases/

ESR is responsible for the surveillance of notifiable diseases in New Zealand and oversees the national notifiable disease database (EpiSurv) on behalf of the Ministry of Health. The surveillance reports encompass analyses and interpretations of data, predominantly from notifiable diseases, influenza, and sexually transmitted infections. The site offers monthly and annual surveillance reports, annual summary tables, and other valuable resources. There is also an interactive dashboard developed by ESR that enhances access to notifiable disease data.

Datasets from international sources

data.world
https://data.world/

data.world is a collaborative platform that hosts a diverse collection of datasets contributed by individuals, organizations, and communities from around the world, fostering data sharing, collaboration, and analysis across various domains and topics.

Keywords: collaborative data, data sharing, data analysis, community-contributed data, global datasets

Google Dataset Search
https://datasetsearch.research.google.com/

Google Dataset Search is a tool that allows users to discover and access datasets from various publishers across the web, enabling researchers, scientists, and analysts to find valuable datasets for their work and exploration. More information here:
https://www.blog.google/products/search/making-it-easier-discover-datasets/

Keywords: dataset search, research datasets, data access, data sources, Google search, research tools

IPUMS International
https://international.ipums.org/international/

IPUMS International is a platform that provides access to integrated census and survey data from various countries around the world, enabling researchers and analysts to perform cross-national and comparative analyses on population and socioeconomic variables.

Keywords: international census data, cross-national analysis, socioeconomic variables, populations

Kaggle
https://www.kaggle.com/

Kaggle is a data science community and platform that offers a wide variety of datasets, competitions, and notebooks for data enthusiasts, analysts, and machine learning practitioners. The platform facilitates data exploration, analysis, and modelling in diverse domains.

Keywords: data science community, datasets, data competitions, machine learning, data analysis, forums, tutorials, Artificial Intelligence

IPU Data
https://data.ipu.org/

The Inter-Parliamentary Union (IPU) Data platform provides access to a wide range of data related to national parliaments and parliamentary systems worldwide. It offers statistical insights, comparative analyses, and visualisations on parliamentary representation, elections, and political participation.

Keywords: parliamentary data, political, national parliaments, elections data, comparative analysis

NHANES (National Health and Nutrition Examination Survey)
https://www.cdc.gov/nchs/nhanes/index.htm

NHANES is an ongoing survey program conducted by the National Centre for Health Statistics (NCHS), collecting comprehensive health and nutritional data from a nationally representative sample of individuals in the United States. The data covers a wide range of health-related topics, including demographics, health conditions, nutrition, and physical examinations.

Keywords: health, nutrition, CDC, health conditions, demographics, physical examinations

Gapminder
https://www.gapminder.org/

Gapminder aims to promote a fact-based worldview by providing a variety of data visualisations and datasets that cover global development and socio-economic trends. Gapminder's tools and resources enable users to explore and understand how various indicators change over time and across countries.

Keywords: global development, visualisation, socio-economic trends, indicators, fact-based worldview

Our World in Data
https://ourworldindata.org/

Our World in Data is a project that presents empirical data and research on various global topics, including health, environment, technology, and more. It provides interactive visualizations and articles to facilitate understanding and informed discussions about the state of the world.

Keywords: empirical research, data visualisation, global issues, world statistics, data-driven insights

IASE (*International Association for Statistical Education*), ISLP (*International Statistical Literacy Project*), and ProCivicStat
http://iase-web.org/
http://iase-web.org/islp/
http://iase-web.org/islp/pcs/

These sites lead to datasets, other resources, and CivicStatMap (also listed in the *'Other Resources'* section). ProCivicStat is a project by Professor Jim Ridgway that promotes civic engagement through statistics teaching, encouraging students to explore and analyse data related to current social issues. Datasets and metadata are available along with information about main variables, their types, and possible crosstabulations.

Keywords: civic engagement, social issues, data visualisation, variables, crosstabulations, analysis

Carnegie Mellon University - ISLE Data Explorers
http://www.stat.cmu.edu/isle/data_explorers.html

The ISLE Data Explorers website by Carnegie Mellon University provides resources for teaching statistics using real-world datasets. The platform offers various datasets, activities, and teaching materials that allow educators and students to explore data, apply statistical concepts, and enhance their understanding of statistics.

Keywords: real-world datasets, data exploration, statistical concepts, education, teaching materials

Robb Pruim's Resources for Teaching with a Spreadsheet

https://rpruim.github.io/stat-lit/Resources.html

Robb Pruim's website offers resources for teaching statistics using spreadsheets. The platform provides a collection of teaching materials, datasets, and activities designed to engage students in statistical thinking and data analysis using spreadsheet tools.

Keywords: spreadsheet tools, statistical thinking, analysis, teaching materials, education, resources

CODAP Data Science Resources

https://concord-consortium.github.io/codap-data/

The CODAP Data Science Resources website provides a collection of datasets, activities, and examples designed to support teaching and learning data science using the CODAP (Common Online Data Analysis Platform) tool. These resources aim to enhance data literacy and analytical skills for educators and students.

Keywords: data science, data analysis, teaching resources, data literacy, educational activities, examples

World Health Organization (WHO) Data

https://www.who.int/data

The World Health Organization's data platform offers a comprehensive collection of global health-related datasets, statistics, and indicators covering a wide range of topics and regions. This includes data on epidemiology, vaccinations and immunisation, healthcare infrastructure, health financing, non-communicable diseases, environmental health, health inequalities, global health indicators, and country-specific data.

Keywords: global health data, health indicators, international health statistics, health-related datasets

OECD Data

https://data.oecd.org/

Description: The OECD Data platform offers a comprehensive view on economic, social, and environmental data from OECD member and partner countries, allowing users to explore and analyse international trends and comparisons. Some of the key topics covered in these datasets include economic indicators, education statistics, labour market data, health expenditure, environmental performance, trade and foreign aid, tax revenue, public governance, business and entrepreneurship, energy statistics, transportation indicators, and more.

Keywords: international data, economic indicators, social indicators, environmental data, member countries

UNICEF Data - Datasets

https://data.unicef.org/resources/resource-type/datasets/

The UNICEF Data platform offers a diverse collection of datasets related to child and maternal health, education, nutrition, water and sanitation, and more, providing valuable insights into global issues affecting children and families.

Keywords: child health, maternal health, education data, nutrition data, global, child welfare, international development

NASA Earthdata

https://www.earthdata.nasa.gov/learn/find-data

NASA Earthdata offers access to a wide range of Earth science datasets, satellite observations, and environmental data that enable researchers, educators, and students to explore and analyse various aspects of our planet.

Keywords: satellite observations, environmental data, global datasets, scientific research, Earth observations

World Bank Data

https://data.worldbank.org/

The World Bank Data platform offers a comprehensive collection of global economic, social, and environmental datasets, allowing users to explore and analyse development indicators and trends across countries.

Keywords: global development data, economic indicators, social indicators, environmental data, international data

UCI Machine Learning Repository

https://archive.ics.uci.edu/datasets

The UCI Machine Learning Repository provides a diverse collection of datasets for machine learning and data mining research, contributed by researchers and institutions globally.

Keywords: UCI Machine Learning Repository, machine learning, data mining, research datasets, international data sources, data science, artificial intelligence

Data.gov

https://www.data.gov/

Data.gov is the home of the U.S. Government's Open Data that offers access to a wide range of datasets from various government agencies, providing valuable information for research, analysis, and transparency.

Keywords: U.S. government datasets, government agencies, research data, data transparency, data access

Other Resources

These sites contain resource for statistics teaching, including some datasets.

CensusAtSchool: Guide to Resources

https://new.censusatschool.org.nz/resource/a-guide-to-recently-created-statistics-resources-for-aotearoa-new-zealand-2022

The CensusAtSchool New Zealand website offers a comprehensive guide to recently created statistics resources tailored for Aotearoa New Zealand in 2022. It provides educators and students with valuable materials and tools for teaching and learning statistics, making it easier to explore data and engage with statistical concepts.

Keywords: statistics resources, education, teaching, learning, data exploration, statistical concepts

iNZight and iNZightLite

https://inzight.nz/
https://lite.docker.stat.auckland.ac.nz/

iNZight and iNZightLite are data analysis and visualisation tools designed for teaching and learning statistics. Developed by the University of Auckland and also Victoria University of Wellington, they provide interactive platforms that allows users to explore datasets, perform various statistical analyses, and create visualisations to enhance the users' ability to uncover and understand messages in data and statistical concepts.
iNZight and iNZightLite contain a set of ready-to-use datasets in csv form.
iNZight is the version for download, and iNZightLite is the online version that students can readily run from a range of devices.

Keywords: data analysis, data visualisation, statistics education, interactive platform, statistical concepts, teaching and learning

BBC News

https://www.bbc.com/news/science-environment-46384067

Description: This BBC News article provides an overview of the current state of climate change using data visualizations and charts. It offers insights into key climate change indicators and suggests actions individuals can take to address the issue. While the article provides insights into the study findings, it is not a dataset itself.

Keywords: climate change, data visualisation, climate indicators, environmental impact, individual actions, global issue

CivicStatMap

https://rstudio.up.pt/shiny/users/pcs/CivicStatMap_English/

This RStudio Shiny app provides an interactive map that likely displays civic-related statistical data, allowing users to explore and visualise geographic patterns of various indicators.

Keywords: RStudio Shiny, interactive map, civic statistics, data visualization, geographic patterns, indicator mapping

MathsLink

https://www.mathslink.org/statistics

MathsLink offers a wealth of comprehensive resources for statistics education, catering to students, educators, and enthusiasts. It provides a wide range of materials, including tutorials, calculators, and practice problems, making it a valuable platform for learning and improving statistical skills.

Keywords: statistics education, tutorials, calculators, practice problems, statistical skills, mathematics, learning platform

Data Education in Schools

https://dataschools.education/

Data Education in Schools is an online platform from Scotland that is dedicated to enhancing data literacy and analytical skills. It offers a variety of resources, including courses, workshops, and tutorials, to help individuals excel in data-related fields, including plenty of New Zealand contents.

Keywords: data literacy, analytical skills, online courses, workshops, tutorials, data professionals, learning opportunities, data-related fields

The Wonderful World of Statistics

https://sites.google.com/view/uoa-stats-399/home

The University of Auckland's STATS 399 video site hosts a collection of videos created by statistics students as part of a STATS 399 assignment. This page hosts a set of videos created by students from STATS 399 at the University of Auckland in Semester 2, 2022. They may be used to help introduce a topic, showcase an interesting example, or provide an additional explanation in statistics classrooms.

Keywords: videos, student explanations, real world examples, context, fun, engagement

The Open Data Initiative

Open research embodies the principles of transparency, openness, verification, and reproducibility in conducting and sharing research. It leverages digital technology to maximise accessibility. When governments make their data accessible and reusable, they empower individuals, organisations, and governments themselves to foster innovation and collaboration. Open data not only accelerates economic growth but also enhances government accountability, benefiting citizens, organisations, and governments alike. Open data is characterised by open licenses, accessibility, and both human-readable and machine-readable formats, making it a versatile and ubiquitous resource in today's information-driven world.

data.govt.nz: Open Data
https://data.govt.nz/toolkit/open-data/

The data.govt.nz Toolkit offers comprehensive guidance on open data principles, helping New Zealand's public sector organizations make data accessible and transparent. This resource equips users with valuable information, tools, and best practices for publishing open data, enabling data-driven decision-making and innovation. With a focus on openness, collaboration, and engagement, the Toolkit promotes the sharing of machine-readable, reusable data to benefit the government, businesses, and the public. It outlines the process of opening data step-by-step, from planning and licensing to publishing and maintaining datasets. By embracing open data practices, public entities in New Zealand can unlock the potential of their data resources, fostering economic growth, innovation, and accountability.

Keywords: open data principles, guides, transparency, collaboration, engagement

Open Data Institute
https://www.theodi.org/

The Open Data Institute (ODI) is a prominent organisation focused on promoting open data and its impact on society. Their website serves as a hub for information, resources, and initiatives related to open data. The ODI offers valuable insights, research, and tools to help individuals and organisations understand and harness the power of open data for various purposes, including transparency, innovation, and economic growth. They produce free tools and guides to help manage, publish and use data but do not appear to host datasets themselves.

Keywords: open data, transparency, innovation, data impact, data resources, data initiatives

Open Knowledge Foundation
https://okfn.org/en/

The Open Knowledge Foundation is a global non-profit organisation that is working towards a fair, sustainable, and open future where all knowledge is accessible to everyone. The foundation operates with the vision of a world that is "open by design." It offers a range of open knowledge data tools including a data management system, open-source toolkit, and a collection of tools and licenses for open data.

Keywords: open-source data management, open-source toolkit, open data licenses, global movement