



Determinants of Equity Data fundamentals

Objective: This document offers an overview and definitions of fundamental concepts essential for interpreting and analyzing data including what is data, data types, data fitness, and how to understand data visualization charts. It also provides an overview on how the data sets of the DoE were identified, and data limitations.

What is data? Data, or “datum” from its Latin origin means “given thing” and in common modern language it can be said that data is a combination of observable information. Data can take the form of visual images, numbers, quotes, statistics, etc. In this context, data represents the raw facts and figures from which useful information can be extracted.

Types of data

There are two main types of data, qualitative (quality) and quantitative (quantity). The majority of the indicators included in the Determinants of Equity (DoE) data express quantitative data but also offers some qualitative data to better inform population specific conditions, root causes of problems and in general, to get a fuller picture of our population.

Data type	Source and examples	Example data in DoE data tool	Measures
Quantitative	US Census; Numbers, percentages, maps	Percent of the population aged 16-24 who were either students or employed (including in the armed forces)	Measures of values or counts and are expressed as numbers. Data about numeric variables (e.g., how many, how much or how often)
Qualitative	Reports from focus groups, testimonies and quotes	Compilation of community engagement input pertaining law and justice in King County	Describes qualities or characteristics. It is collected using questionnaires, interviews, or observations, and frequently appears in narrative form.

Data limitations

The Determinants of Equity data tool draws heavily on federal, state, and local data. Not all data sources disaggregate race and ethnicity data, and important dynamics and differences in outcomes can be lost when groups are combined into broad categories, missing the intersectionality inherent in human experience. For example, the federal government recognizes there is a systemic lack of data disaggregation in Asian American and Native Hawaiian and Pacific Islander communities. Disaggregated data can unmask realities and needs within Asian Americans and Pacific Islanders subgroups, leading to better policies.

Data justice: The Determinants of Equity works under a data justice framework. Data justice aims to capture forms of knowledge and lived experiences that are community-centered and community-driven to counter the systemic erasure and harm perpetrated on BIPOC communities via oppressive data practices. The fundamental premises of data justice are that data should: (1) make visible community-

Sources:

1. Brown, M. Transforming Unstructured Data into Useful Information
2. [US Census. Gov- American community survey data suppression](#)
3. The Coalition of Communities of Color

driven needs, challenges, and strengths, (2) be representative of community; and (3) treat data in ways that promote community self-determination. For example, the DoE data tool offers insight into community priorities and solutions to help contextualize the qualitative data offered in the data charts.

Interpreting and analyzing data

Data analysis: Data means nothing by itself-- data analysis is the process of inspecting, cleansing, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.

Technical notes: As with other data tools, the DoE uses and displays technical notes that contain details on how data was gathered and analyzed, short explanations of concepts, and descriptions on a modification of an existing data collection techniques or procedures. These types of technical notes are the most frequently used in the DoE data tool:

Confidence intervals: A confidence interval is a range around an estimate that conveys how precise the measurement is. Much of the DoE data are estimates and not exact counts. The confidence interval is the range around an estimate where statistically the precise count exists. For most of the data included in the DoE data tool, margin of error presented is 95. Confidence intervals are calculated based on the standard error of a measurement. For sample surveys, such as the presidential telephone poll, the standard error is a calculation which shows how well the poll (sample point estimate) can be used to approximate the true value (population parameter), i.e., how many of the people surveyed responded yes to a specific question. For instance, suppose you want to estimate the average height of all adult males in a city. Measuring every single adult male in the city is impractical, so you take a random sample of 100 adult males and measure their heights. Having measured your sample you find that the sample mean is 175cm. You also want to calculate the standard deviation to create an estimate that fits your entire male population. In doing so, you will be to be 95% confident that the most approximate average height of all adult males in the city lies between 173.04 cm and 176.96 cm.

Rolling averages: When the frequency of an event varies widely from year to year, rates are sometimes aggregated into averages, often into three-years interval to smooth out the peaks and valleys of the yearly data. For example, for events occurring from 2001 to 2017, rates may be graphed with the following rolling averages: 2001-2003, 2002- 2004....2015-2017.

Data suppression: To limit the disclosure of information about individual respondents and to reduce the number of estimates with unacceptable levels of statistical reliability, specific estimates may be suppressed or not provided by the data source. These values are marked as suppressed in the dataset.

Uses of the DoE data tool: When agencies are preparing to develop a new program or initiative, or to fund services, it is important to understand the baseline measurements of the policy domain the project is intended to impact. This allows for tracking change over time and see whether the intended outcomes were reached. For example, the DoE data tool is a great starting place to find the baseline measurement for your performance measures.

Sources:

1. Brown, M. Transforming Unstructured Data into Useful Information
2. [US Census. Gov- American community survey data suppression](#)
3. The Coalition of Communities of Color

Data fitness

Data fitness: The following criteria provide specifications that were taken into consideration when selecting indicators to include in the DoE data tool. This criterion provides assurance that data sets included are consistent, usable, and able to inform required and desired county functions, including decision making, evaluation, planning, and comparisons:

Acceptability	Are data sets sufficient to be able to use and trust the data for decision making, planning, implementation, etc.?
Completeness	Minimum of threshold for missing data across all data sets.
Availability	Population level data departments collect in a yearly/ cyclical basis
Accuracy	Form and content: are the data reliable, and correct? right / true values, represented consistently across sample.
Reliability	Complete and accurate, trust of data source
Timeliness	Does data represent a real and timely picture of current county situation? Does it show trends over time?
Comparability	Metadata and data content that is consistent across datasets
Repeatability	Likelihood that data measures can produce same results when replicated

How to understand data visualization charts

Reading a data chart: Below is a data chart from the DoE data tool with the various components of information in the chart. Understanding these components is key in being able to data analysis.

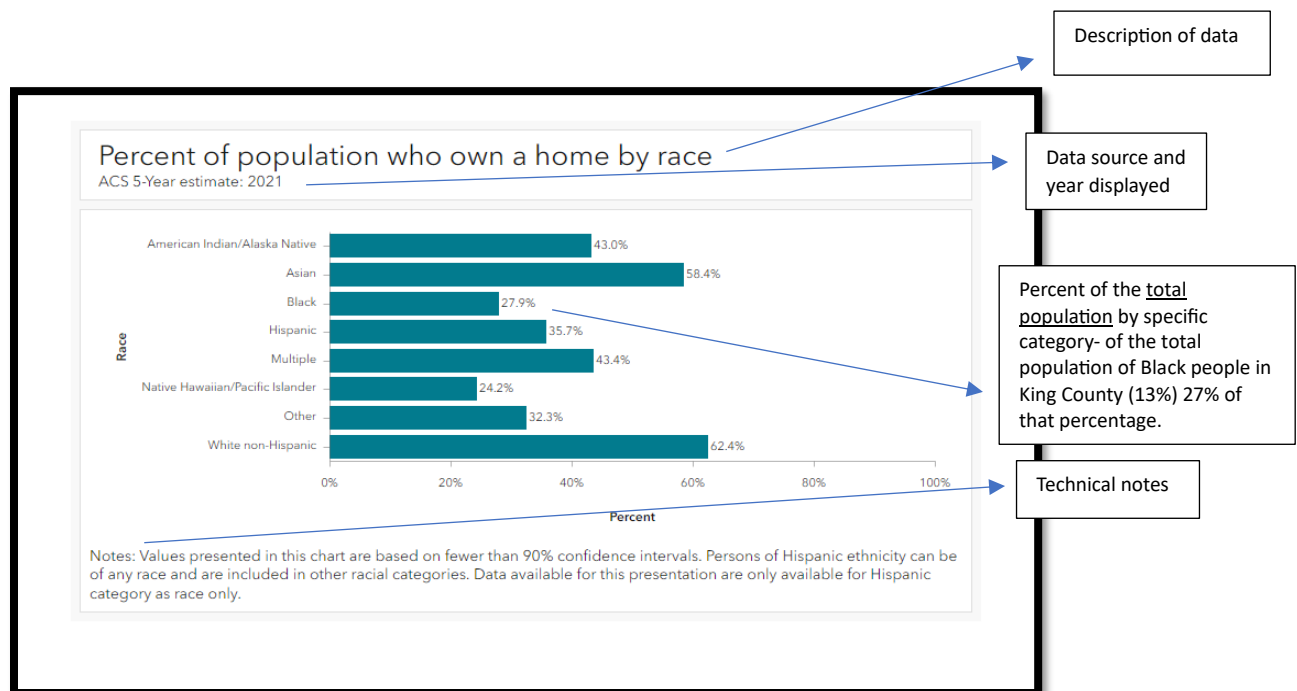


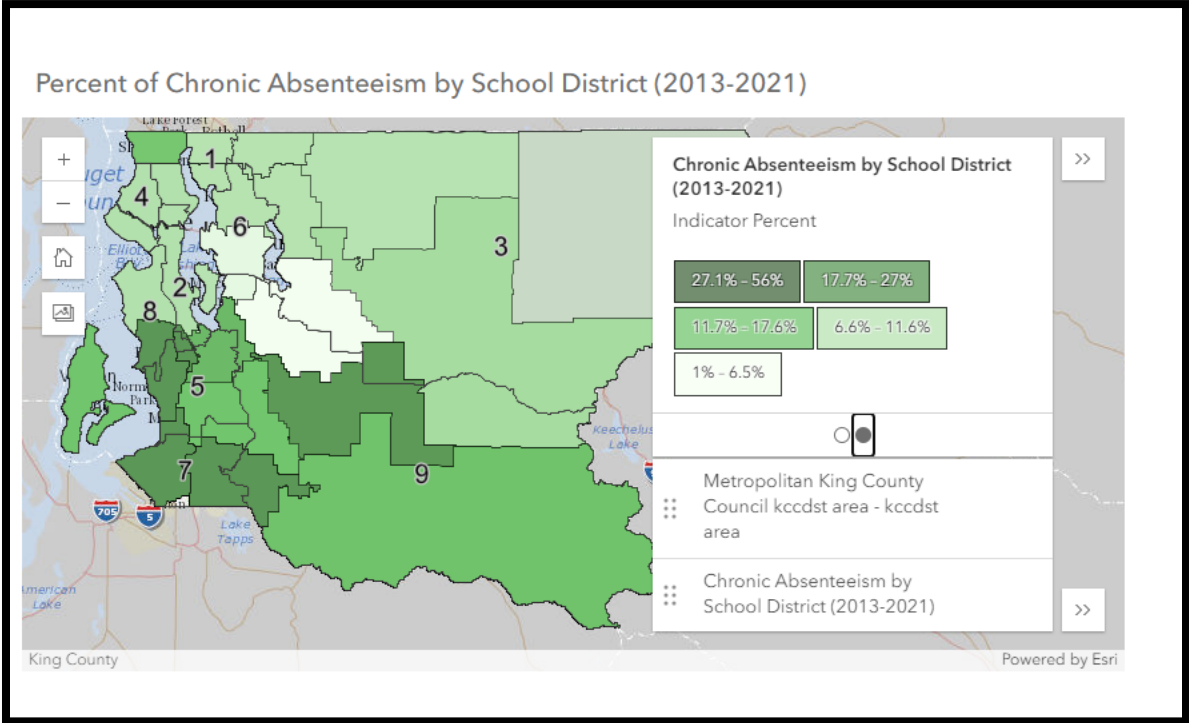
Chart Component	Description
-----------------	-------------

Sources:

1. Brown, M. Transforming Unstructured Data into Useful Information
2. [US Census. Gov- American community survey data suppression](#)
3. The Coalition of Communities of Color

Title	The title describes the data presented in the chart. In this example, this chart describes how homeownership (people living in a home their family owns) varies by their race.
Source	The data used in this chart comes from the U.S. Census Bureau’s American Community Survey (ACS) data, specifically the 5-year data series for the year 2021. The ACS is a continuously run survey that aggregates responses over a single year or over 5 years and reports on demographics of communities across the United States. While the 1-year series estimates allow for annual comparisons of larger cities and counties, the 5-year series data reports detailed demographics for smaller places (cities, towns, ZIP codes) or customizable areas. Because the 5-year series includes overlapping years of data (e.g., 2021 includes 2017-2021 and 2022 includes 2018-2022) users should only compare non-overlapping series of data (e.g., 2011-2015 vs. 2017-2021).
Bar Chart	We know from the title that this chart is describing the relative shares of homeownership for people grouped by their racial identity. The vertical Y-axis shows the different racial categories reported for this demographic. The horizontal, X-axis displays a percentage scale ranging from zero to 100 percent. The data bars show the intersection of these concepts: the percentage of the population within each racial category that lives in a home their household owns.

Reading a data geographical map: Below is a geographical map from one of the Determinants of Equity data tools. Explanation of its components is detailed in the table below it. Understanding these components is essential for equity impact analysis.



Map component	Description
---------------	-------------

- Sources:
1. Brown, M. Transforming Unstructured Data into Useful Information
 2. [US Census. Gov- American community survey data suppression](#)
 3. The Coalition of Communities of Color

Geographical area (map extent)	The map focuses on the school districts within King County. This map provides a visual representation of chronic absenteeism across different school districts, helping stakeholders identify areas with higher or lower absenteeism rates.
Numbers	The map is divided into numbered regions, each of which represents a different Metropolitan King County Council area and school district.
Legend	The right-hand side of these maps contains a legend explaining the color coding of the districts.
Color coding	The map provides a different variation of shades to indicate varying levels of chronic absenteeism in each school district. The colors range from high incidence of absenteeism (dark green) to lower incidence of absenteeism (very light green). So for example, Districts shaded in darker green (e.g., District 4 and District 5) have higher percentages of chronic absenteeism, ranging from 27.1% to 56%. Districts shaded in lighter green (e.g., District 1 and District 2) have lower percentages of chronic absenteeism, ranging from 1% to 6.5%.

Sources:

1. Brown, M. Transforming Unstructured Data into Useful Information
2. [US Census. Gov- American community survey data suppression](#)
3. The Coalition of Communities of Color