Outline

* About the data set
  + Rows and columns
* How data visualization helps derive meaning from data
  + Interpretation of plots
  + Importance to of looking at the axis units and range
  + Calculate the delta for looking at trends
* How measurements were taken­ and how it would influence results
* How do Canadians stack up compared with fitness standards
  + Effect of age vs. fitness level

Previous titles:

* Data Visualization of Statistics Canada's Physical Fitness Measures of the Household Population

# Data Visualization of Canada's Fitness as Measured by Stats Canada

When was the last time you looked at a large table of numbers and thought, "Wow, it's so obvious what's going on." Unless you had a compelling reason to squint at each line and column (e.g., work or school), it's not likely that a spreadsheet meant much of anything at first (and likely only) glance. Here's what this article will cover:

1. An example of how data can be wrangled and visualized for easier interpretation.
2. What you should pay attention to when looking at visual representations of data, whether you are creating the graphs or just interpreting them.
3. How are Canadians doing in terms of fitness, according to Statistics Canada data?

About the data

The Physical fitness measures of the household population data set (1) from Statistics Canada is from the Canadian Health Measures Survey (CHMS), which "aims to collect important health information through a household interview and direct physical measures at a mobile examination centre" (2). Every two years, various health metrics are measured in a sample representative of Canada's population. Here's a brief overview of the data:

\* Year: The survey has been done every 2 years from 2009 to 2019 (though the data are incomplete in some cases).

\* Demographic information: Age group, and sex.

\* Fitness measurements: Maximum oxygen consumption (VO2max), total grip strength, sit and reach, and partial curl-up were taken.

\* Summary statics: The mean, percentile distribution (5th, 10th, 25th, 50th, 75th, 90th, and 95th  percentiles), and their respective 95% confidence intervals.

The Statistics Canada website provides graphs of some of their data, but I couldn't find any for this particular data set. Regardless, being able to display the data the way you want allows you to make better sense of all the numbers. Once you set up your code in a language such as Python, you can pretty much re-run the code to recreate the graphs as new data become available.

Data Wrangling and Visualization

I've focused on the maximum oxygen consumption (VO2max) and total grip strength measurements for my analyses because I consider them to be the important ones (the sit and reach test has actually been discontinued as fitness guidelines have evolved):

\* VO2max is an indicator of cardiorespiratory fitness that is estimated using some sort of aerobic exercise test such as walking/running. It represents your capacity to do physical work: The higher your VO2max, the less easily you will fatigue during sustained/repetitive activity.

\* Grip strength is measured by asking the participate to squeeze a measuring device as hard as possible with each hand. Having a higher grip strength is usually associated with having greater overall strength, so it's a convenient non-invasive way to gauge strength (practicality is especially important when testing a large sample). Strength is important for performing activities of daily living and reducing fall risk. It is also a proxy for muscle mass, an underappreciated yet important factor for minimizing risk for cardiometabolic diseases such as diabetes and cardiovascular disease.

As a fitness professional with an interest in health promotion, the question I was most interested was whether there is a trend in physical fitness: Which groups are getting fitter? Which groups are getting less fit?

# Cardiorespiratory fitness

Since I was primarily interested in the effect of time and age, this was one of my initial graphs:

[crf vs year by age.png]

Figure 1: Estimated maximal oxygen consumption by over time for each age group (age groups are listed by the lowest age in the age group).

The y-axes are identical in all plots, showing that age and sex have some influence on VO2max (which is already normalized to bodyweight). So far, the only clear time-based trend is that VO2max of the least fit males aged 12-19 have been declining.

To more easily focus on the \*change\* (∆) of values over time for time series data with different starting points, you can simply subtract the value of the first data point (time = 0) from all values:

[crf change vs year by age.png]

Figure 2: Change in estimated maximal oxygen consumption compared with 2009. As seen in subplot d, VO2max values in males aged 12-19 have been declining in the lower percentile groups. Percentile group is based on the absolute values (as shown in Figure 1)

Now that all values at the first time point is zero and the y-axis range is narrower, it's easier to see how values change over time. While many of the lines oscillate, a trend towards declining VO2max can be observed in:

\* Males aged 20-39-year-old in the 25th-75th percentile (subplot f),

\* Males aged 40-59 in the 25th percentile (subplot h), and

\* Males aged 60+ in the 10th percentile (subplot j).

There is a caveat to rescaling the y-axis like this: It may make changes appear more significant than they are. For example, let's just rescale the bottom right plot (60+ females):

[crf change vs year females age 60.png]

Figure 3: The y-axis scale can impact the perceived magnitude of differences in data points. Both subplots are identical except for the y-axis, but the year-to-year changes appear greater in subplot b due to different scaling of the y-axis.

Now the changes from one time point to the next seem more dramatic. However, plotting the delta values same y-axis scale as the absolute values shows that these changes over time are pretty small compared with the variation within the sample:

[crf absolute and change vs year females same yaxis.png]

Figure 4: Subplot a in this figure shows the same data as in Figure 3, but with the same y-axis as the plot of absolute VO2max values.

Indeed, the year-to-year changes are often within the 95% confidence intervals (not shown), meaning the differences are not statistically significant.

There are other ways to show the data depending on which relationships are of interest to us:

[crf change vs year by percentile.png]

Figure 5: Visualization of how values change over time based on relative fitness level.

[crf vs percentile group by age.png]

Figure 6: Visualization of how values change across relative fitness levels for each age group. Aside from males under age 20, there are not noticeable trends across the years.

While there has not been an obvious population-wide improvement or decline in cardiorespiratory fitness since 2009 based on estimated VO2max values, let's see to what extent age and fitness level have an impact in adults.

[crf vs age.png]

Figure 7: VO2max vs. age. Values are means across all years in which data are available with the 95% confidence intervals. Note that the lower body average bodyweight of children partly contributes to the higher VO2max values, since values are normalized to body weight.

I'll come back to this figure later on.

Grip strength

When we visualize the total grip strength data, it seems that females under age 40 have been increasing their grip strength since 2011 (yay!), as seen in subplots a and c:

[grip change vs. year by age.png]

Figure 8: Change in total grip strength compared with 2009.

Similar to the VO2max data shown in Figure 7, Figure 9 shows that grip strength is negatively associated with age among adults. There is also a greater difference between males and females in terms of the average and range.

[grip vs age.png]

Figure 9: Grip strength vs. age in adults. Values are means across all years in which data are available with the 95% confidence intervals. The range of values within age groups far exceeds the range between age groups.

Making sense of the data

How you create your data visualizations requires thought about what it is that you specifically want to understand and communicate. While statistical tests help detect statistical significance, visualizations help us understand the story behind the data. They are also powerful communication tools for creating change.

We can refer to fitness standards to see how Canadians stack up (3). But what I believe is more important and relevant from this data is that getting older doesn't necessarily mean that fitness will decline: There are older individuals with greater fitness levels than younger individuals, particularly for grip strength, as seen in Figure 1. I want to emphasize this because society's general acceptance that that health and fitness decline with age often hinders us from taking the actions that keep us feeling and functioning well as we get older.

It will be interesting to see the how Canadians' fitness levels evolve over the decades. My hope is that more and more people practice the lifestyle habits that help them defy time: moving often, eating healthily, and being happy.

# Bibliography

1. **Statistics Canada.** Physical fitness measures of the household population. *Statistics Canada.* [Online] https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310032401.

2. **Canadian Health Measures Survey (CHMS). *Statistics Canada.* [Online] https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5071#a1.**

**3. *Updated Reference Standards for Cardiorespiratory Fitness Measured with Cardiopulmonary Exercise Testing: Data from the Fitness Registry and the Importance of Exercise National Database (FRIEND).* Kaminsky, L., et al. 2, s.l. : Mayo Clinic Proceedings, 2022, Vol. 97, pp. 285-293.**