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

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

When was the last time 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. Example of how a data set can be wrangled and visualized for easier interpretation.
2. What you should pay attention to when looking at graphical representations of data, whether you are creating the graphs or just interpreting them.
3. What do the data tell us about Canadians' fitness?

# 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 of Canadians that is intended to be representative of Canada's population. Here's a brief overview of the data:

* Year. There is data every year from 2009 to 2019 (though the data are incomplete in some cases).
* Demographic information: Age group, and sex.
* Fitness measurement taken: maximum oxygen consumption (VO2max), total grip strength, sit and reach, and partial curl-up.
* Summary statics: The mean, percentile distribution (5th, 10th, 25th, 50th, 75th, 90th, 95th  percentiles), and their respective 95% confidence intervals.

The Statistics Canada website does provide 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 with updated fitness guidelines). [explain why these 2 indicators are important]

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 group 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]

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 males under age 20 have been declining.

To more easily focus on the \*change\* on 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]

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, some additional trends become evident when you look at the 10 year range: Values have been declining for males in the 40-59 and 60+ age groups in the 25th percentile, and in the 20-39-year-old age groups in the 25th-75th percentile.

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]

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]

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

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

[crf change vs year by percentile.png] Caption: Visualization of how values change over time based on relative fitness level.

[crf vs percentile group by age.png] Caption: Caption: Visualization of how values change across relative fitness levels for each age group. Aside from males under age 20, there are not noticeable variations across the years.

While there has not been an obvious population-wide trend 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] Caption: 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 bodyweight.

## 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!).

[grip change vs. year by age.png]

Similar to with VO2max, grip strength also shows that grip strength is negatively associated with age among adults.

[grip vs age.png]

Figure 1: 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.**