

LOOKING TO THE PAST TO PREDICT THE FUTURE

Today, information is **abundant** yet often **complex**.

Statisticians play a **pivotal role** in harnessing data to unravel patterns, trends, and insights in numerous domains to **extract meaningful predictions**.

Here, **American Statistical Association members** with expertise in climate science, sports analytics, finance, and epidemiology share public data sets, data visualization tools, and key factors to consider when making predictions in their areas.



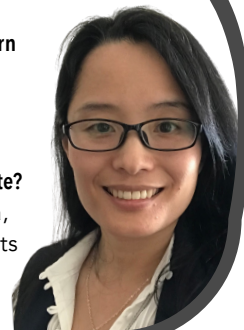
BO LI, UNIVERSITY OF ILLINOIS

What data set would you suggest for high-school students and their teachers if they want to learn about how predictions are made in your area?

National Oceanic and Atmospheric Administration Climate Data Online (www.ncei.noaa.gov/cdo-web)

What are some key factors you think about when developing a model to make predictions about climate?

Exploratory data analysis to understand the pattern and characteristics of the data, going after adequate but parsimonious models, model diagnostics, verifying results and comparing to baseline models, and uncertainty quantification for predictions.



RON YURKO, CARNEGIE MELLON UNIVERSITY

What data set would you suggest for high-school students and their teachers if they want to learn about how predictions are made in your area?

The SCORE Network data repository (data.scorenetwork.org) and CRAN Task View: Sports Analytics (cran.r-project.org/web/views/SportsAnalytics.html)

What is one prediction you are willing to make about a future event?

Without building a predictive model of any kind, I am extremely confident that American tennis star Coco Gauff will win at least another Grand Slam tournament in her career.



MICHAEL JACKSON, RICE UNIVERSITY

What data set would you suggest for high-school students and their teachers if they want to learn about how predictions are made in your area?

S&P 500 data (bit.ly/SandP500Data) and Wharton Research Data Services (wrds-www.wharton.upenn.edu)

What are some key factors you think about when developing a model to make predictions about finance?

I think the assumptions are a key factor, and determining if the model's assumption fits the finance data. After the assumptions, I think about the amount of data required for the method. After understanding the amount of data required, I determine if I want accuracy or interpretability.



ALEXANDRA HANLON & TANNER BARBOUR, VIRGINIA TECH

What data set would you suggest for high-school students and their teachers if they want to learn about how predictions are made in your area?

The Surveillance, Epidemiology, and End Results (SEER) database (<https://seer.cancer.gov>) and UK Biobank (www.ukbiobank.ac.uk)

What is your favorite tool for data visualization?

My team tends to use R because it is a powerful open-source programming language designed specifically for statistical computing and graphics. It offers a wide array of libraries for high-quality data visualization—including ggplot2, plotly, and ggvis.

