COMPUTATIONAL APPLICATIONS TO POLICY AND STRATEGY (CAPS)

Session 5 – Introduction to Statistical Learning and Applications to StarCraft II

Henry Fung

Outline

1. (Very) Brief History of AI Development (5 min)
2. Introduction to Statistical Learning (30 min)
3. Exploratory Data Analysis in Starcraft (20 min)

Rule based vs neural network approach (1980s)

- AI is a technology that can perform tasks at a human (or super-human) level. But the tasks that it performs is restricted to a very narrow domain (ex: play chess, play Go, facial recognition)
- Rule based approach: if X then Y; hardcode human strategies into the computer.
- Neutral networks approach: mimic how the human brain works with artificial neurons. Feed a lot of data to the computer and let it figure out patterns within the data.

The resurgence of the neural network approach (2000s)

> For decades, the field was dominated by the rule-based approach. This is mainly because of the lack of computational power and data; thus, neural-network programs cannot demonstrate its effectiveness.

> **Geoffrey Hinton:** discovered a new method to train layers of neural networks in a very efficient way. This method is called “deep learning”.

The age of implementation (2010s and onwards)

> **Lee Kai Fu:** The “hard work” of AI research is finished for now.

> While we wait for the next break-through, let’s get down to the dirty work of turning AI algorithms to sustainable businesses.

> **Analogy:** We have electricity: let’s harness it to make microwave ovens, industrial equipment, air-conditioners, and refrigerators.

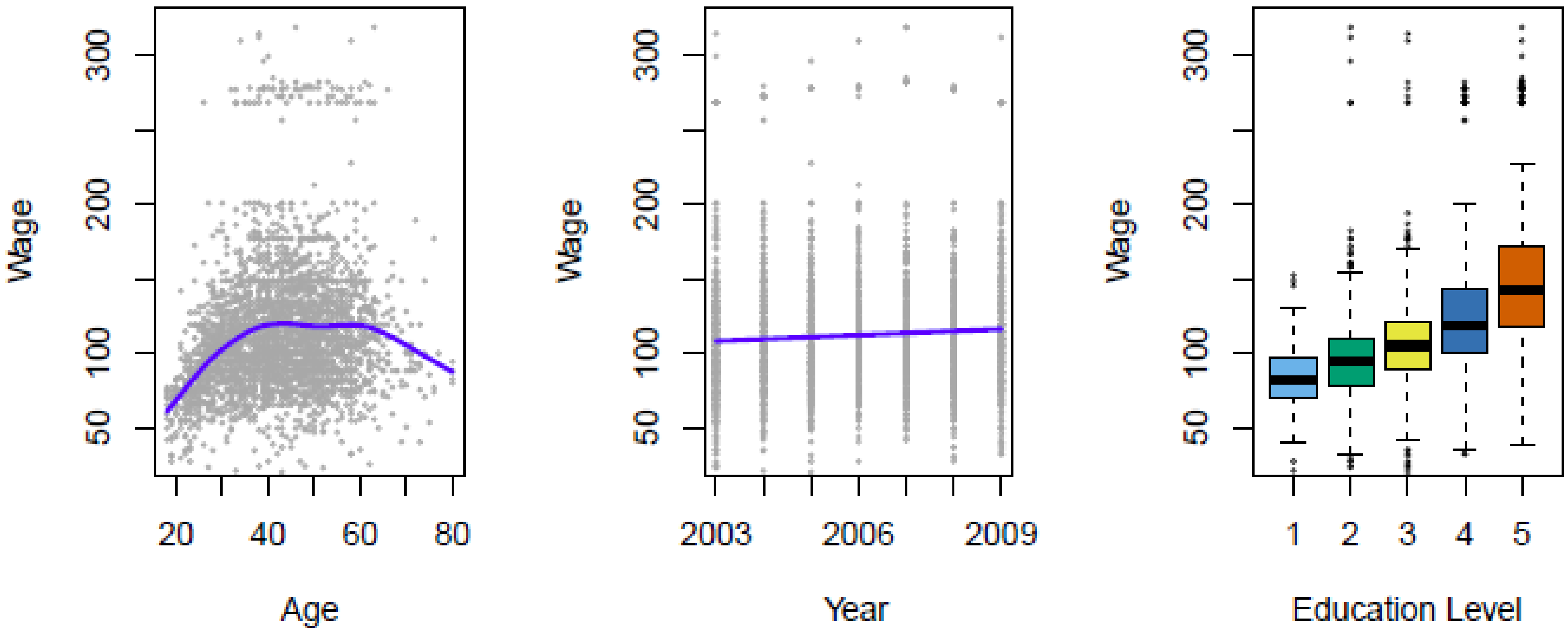
Overview of Statistical Learning

- Statistical learning (aka Machine Learning): a vast set of tools for understanding data.
- These “tools” can be classified as **supervised** and **unsupervised.**
- Supervised Learning:
  - Build a statistical model to *predict* output based on one of more inputs (prediction).
  - Build a statistical model to *infer* the relationship between output and inputs (inference).
- Unsupervised Learning:
  - Learn relationships and structure from data with no supervising output

Example 1: Wage Data

- Dependent Variable/Response Variable:
  - **wage**
- Independent Variable/Feature/Predictor/Regressor:
  - **education, age, year**
- **Objective:** Use supervised learning approach to understand the relationship between wage and age, education and year.
- **Methods:** Linear Regression (and its variants), Generalized Additive Models, Regression Splines

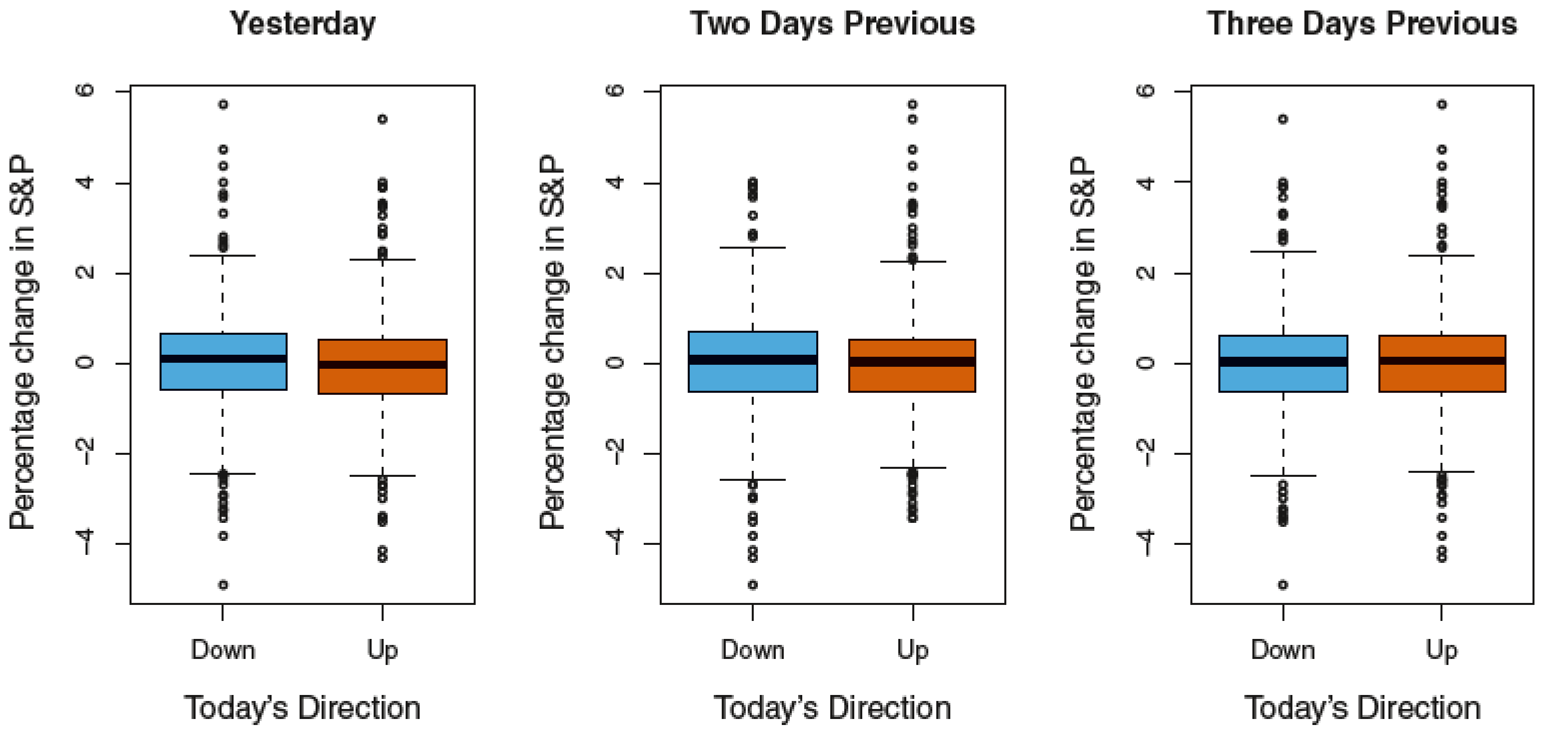
Exploratory Analysis: Wage Data



Example 2: Stock Market Data (Classification Problem)

- Wage data involves a continuous (quantitative) output. Then we have a **regression problem**
  - Other examples: temperature, height, GDP.
- If the data has a discrete (qualitative) output, then we refer this as a **classification problem**
- Stock market data:
  - Output: Direction of the Market (Up/Down)
  - Input: Percent change of yesterday’s S&P stock index
- **Methods:** Logistic Regression, KNN Classifiers

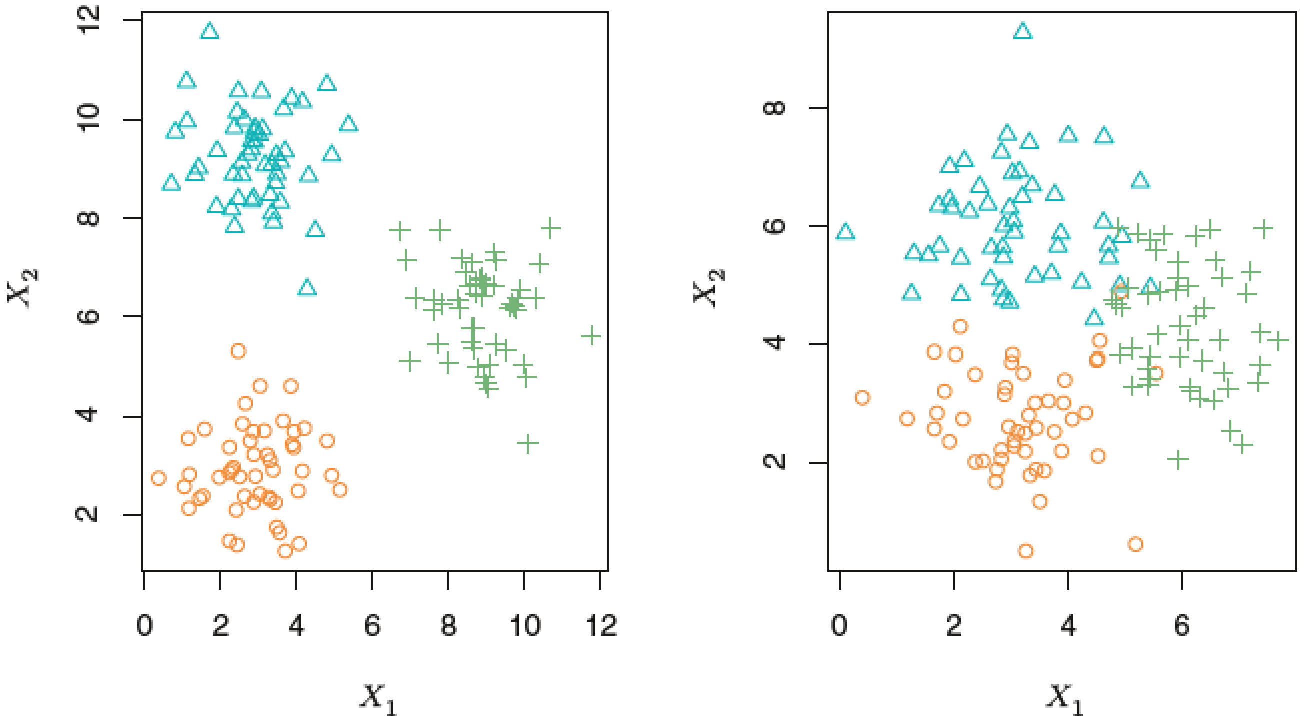
Exploratory Analysis: Stock Market Data



The Clustering problem

- Data with only observed input variables, with no corresponding outputs.
- For example, suppose we collect the demographic data for a number of potential/current customers:
  - Age
  - Spending habits
  - What kind of cars that they drive
  - Whether they rent or own houses
- **The Clustering Problem**: Unlike Example 1, we don’t have information on the individuals’ wage. So we are not trying to predict an output variable. Rather, we want to group individuals according to their observed characteristics

The Clustering problem plot



Interim Review

- History of AI Development
- Introduction Statistical Learning:
  - Supervised vs Unsupervised learning
  - The Regression problem (Wage data)
  - The Classification Problem (Stock Market data)
  - The Clustering Problem (an example of unsupervised learning)
- Next:
  - **Parametric vs Non-parametric models**
  - **The Prediction Accuracy and Model Interpretability Trade-off**
  - **Assessing Model Accuracy and the bias-variance Trade-off**

The “true” population model f(x)

> Y: response variable

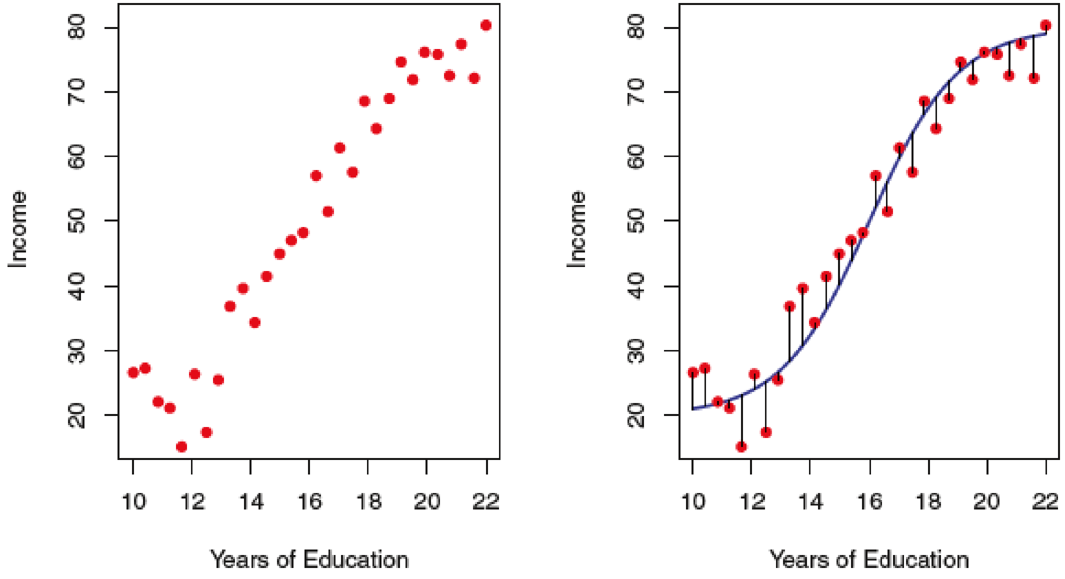
> X: $X_1, X_2, \ldots, X_p.$) predictors

$$Y = f(X) + \epsilon.$$

> The red dots are the observed data

> The blue line is the true function that relates education to income. It is generally unknown in real life.

> The “straight lines” are the errors.



Reducible vs Irreducible errors

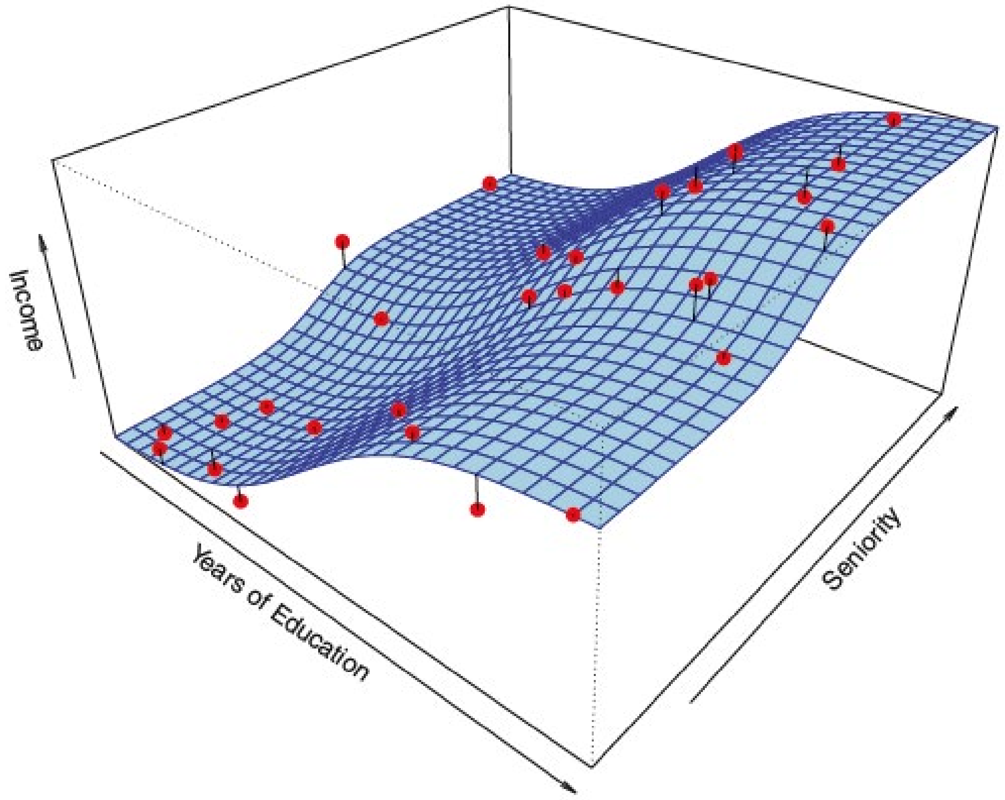
> “True” population model:

$$Y = f(X) + \epsilon.$$

> Prediction:

$$\hat{Y} = \hat{f}(X),$$

> **Lesson # 1:** Even if we can somehow make a “perfect” estimate of f(x) using the data, the accuracy of our Y-hat is still limited by the irreducible error.



Parametric Methods

- **Response variable:** Y
- **Predictors:** P predictors
- **Number of observations in our training data:** n
- **Step 1:** Make an assumption about the functional form of f(x)

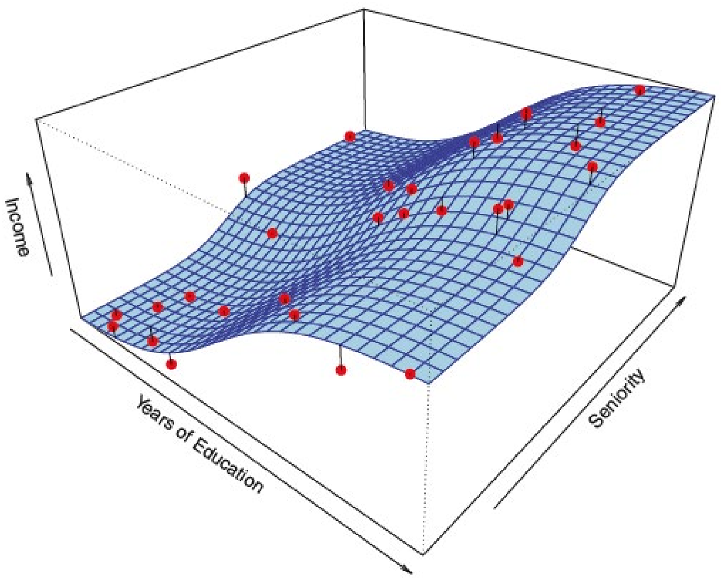
$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p.$$

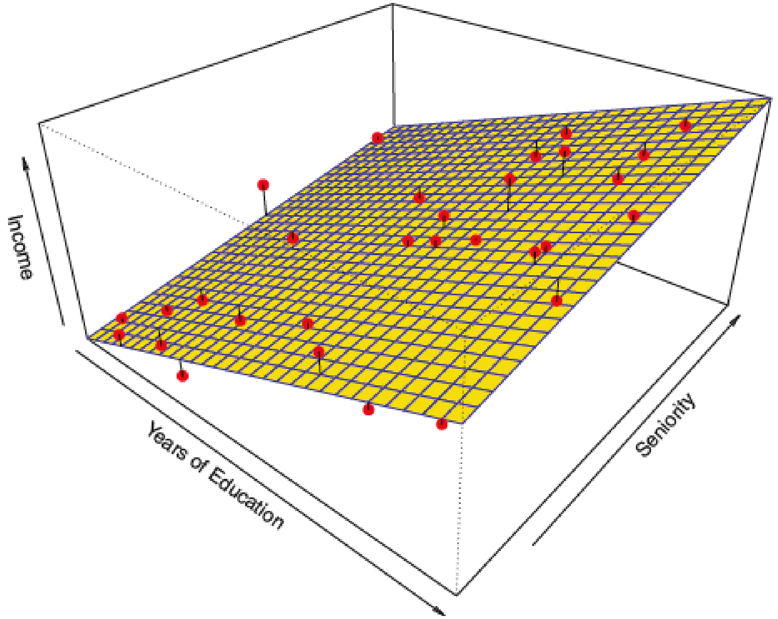
- **Step 2:** Use a procedure (ex: OLS) that uses the training data (our n observations) to “train” the model and estimate the parameters (B’s)

The Pros and Cons of Parametric Methods

> Suppose our estimated model is:

$$\texttt{income} \approx \beta_0 + \beta_1 \times \texttt{education} + \beta_2 \times \texttt{seniority}.$$

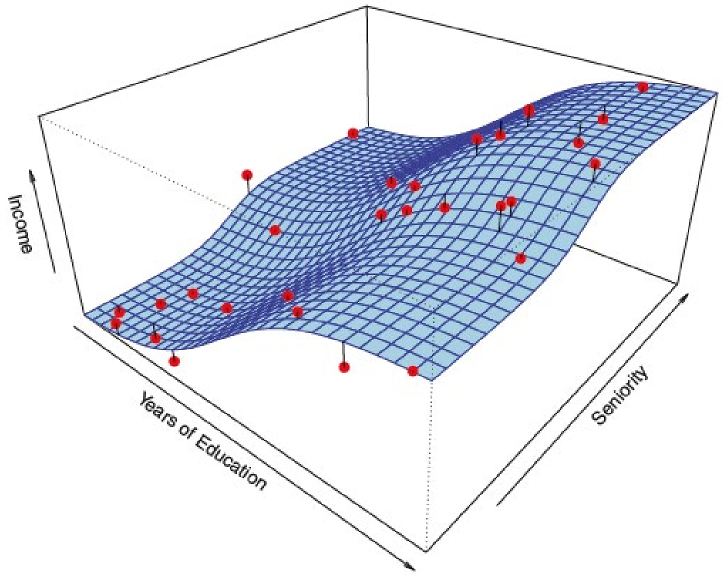


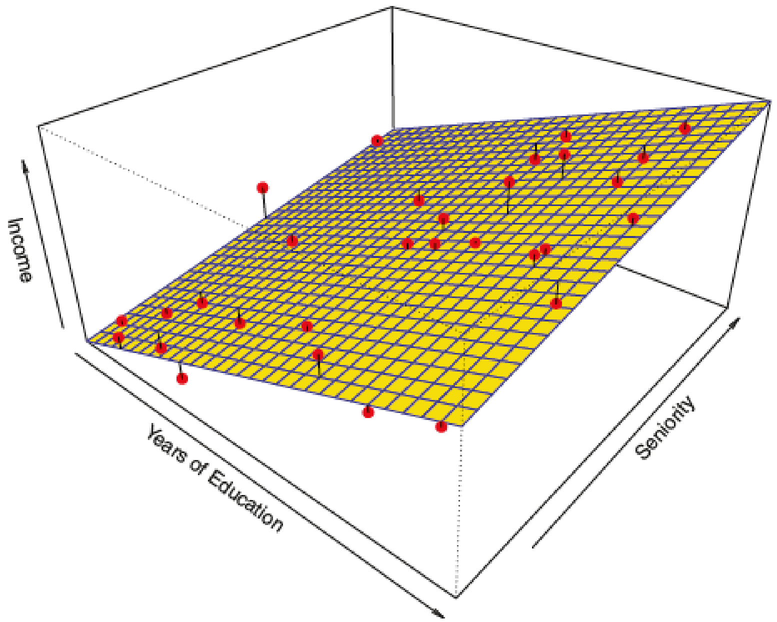


Non-Parametric Methods

> Do not make explicit assumptions about the functional form of f(x).

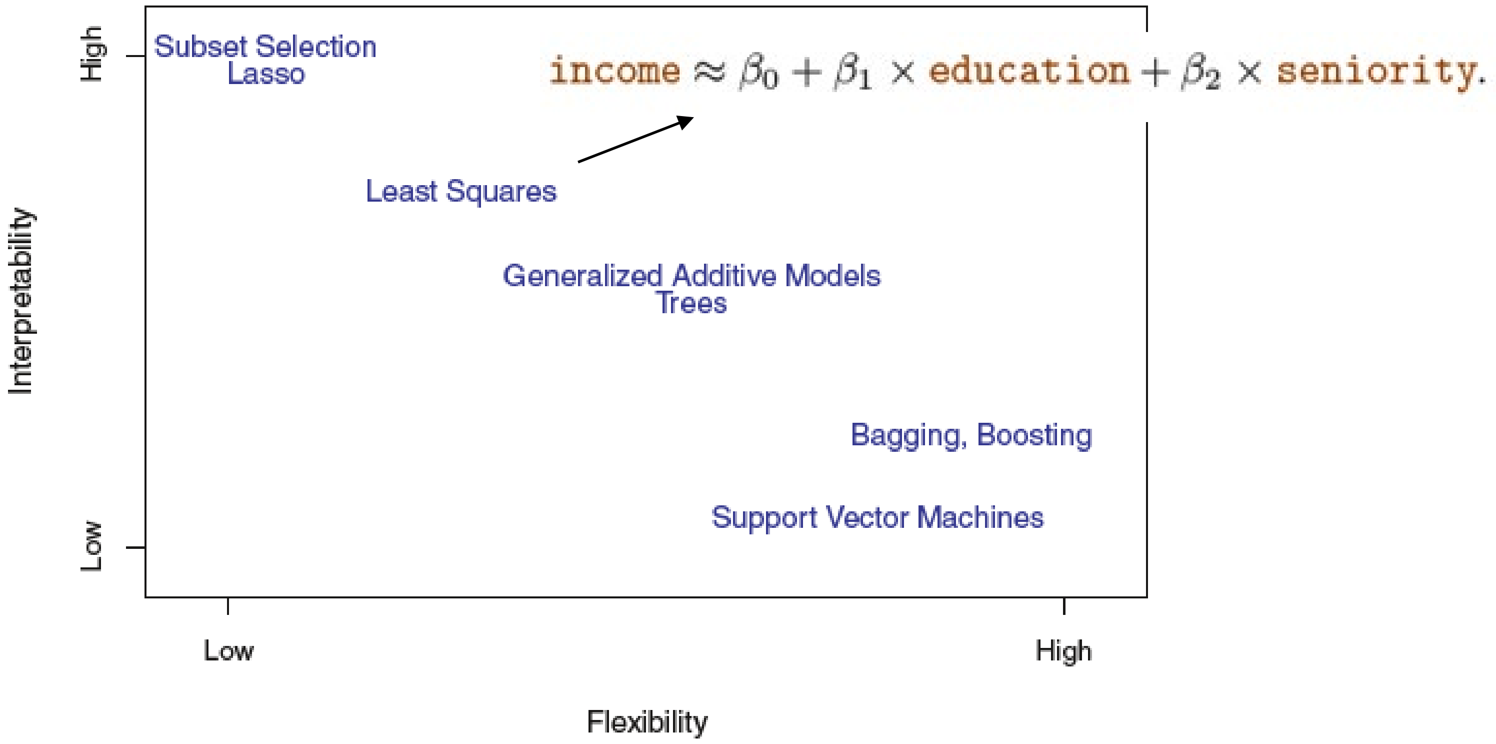
> Only seek an estimate of f(x) that gets as close to the data points as possible without being too “rough” or “wiggly” (formally, without having too much degrees of freedom)





Prediction accuracy vs Model Interpretability

> **Lesson 2:** As the flexibility of a method increase, its interpretability decreases



$$\texttt{income} \approx \beta_0 + \beta_1 \times \texttt{education} + \beta_2 \times \texttt{seniority}.$$

Measuring the Quality of Fit

> To assess model accuracy, the most commonly used measure in the regression problem (besides R-sq) is mean square error:

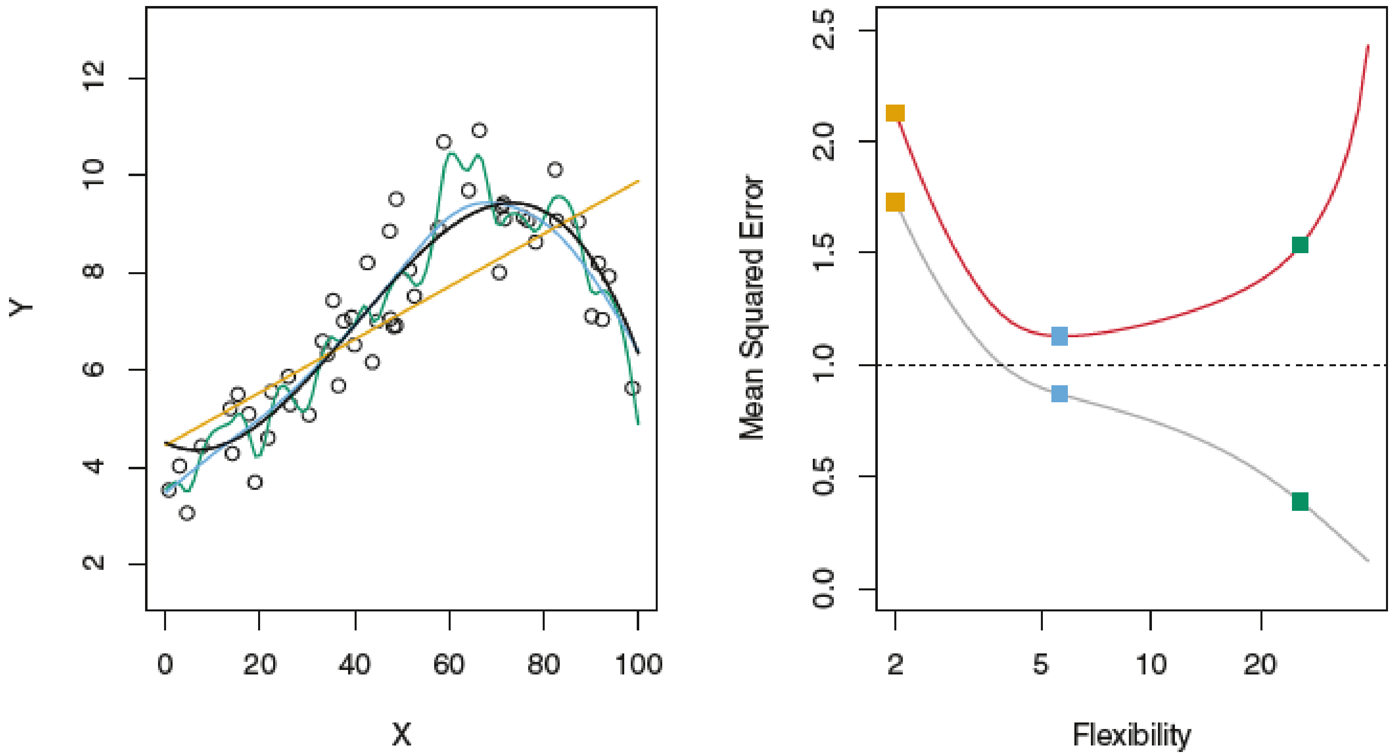
$$Y = f(X) + \epsilon.$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2,$$

$$\text{Ave}(y_0 - \hat{f}(x_0))^2 :$$

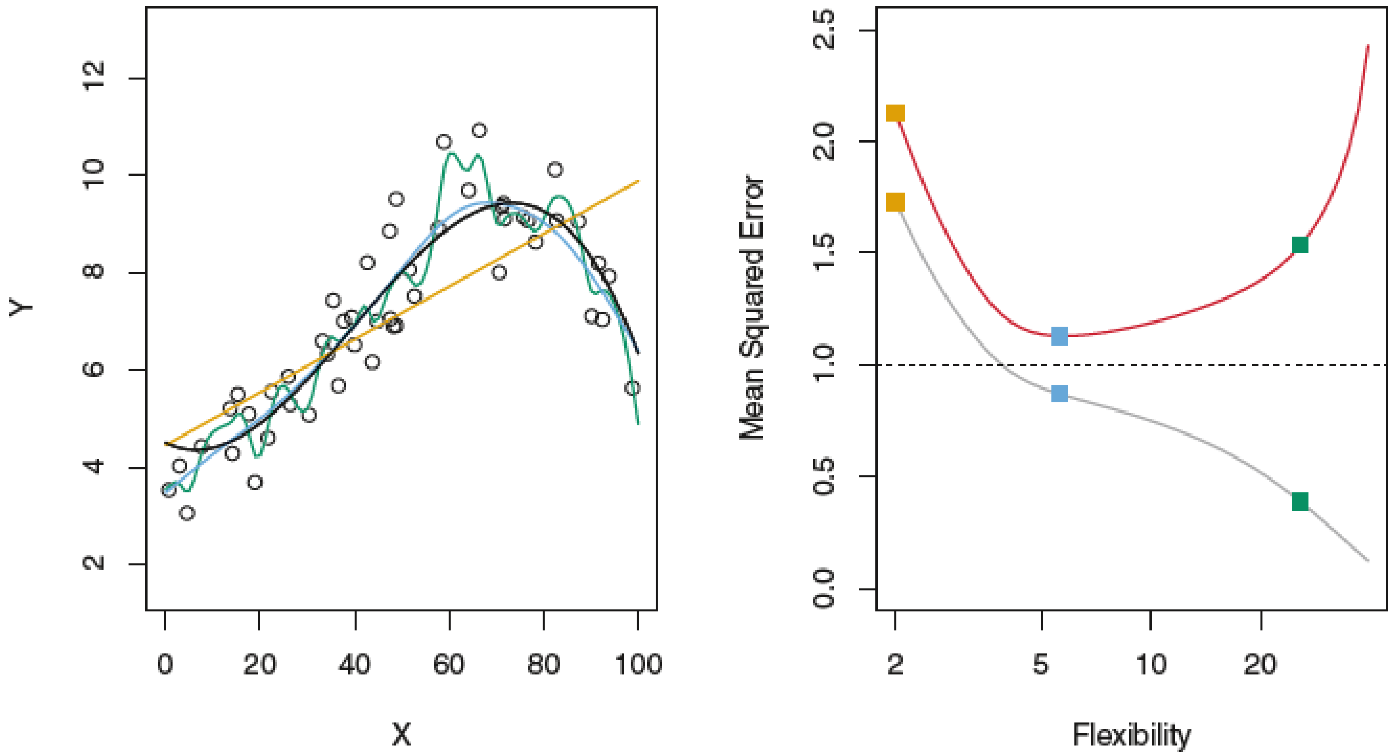
The Bias-Variance Trade-off I

> To minimize the test data MSE, the statistically learning method must simultaneously achieve low variance and low bias

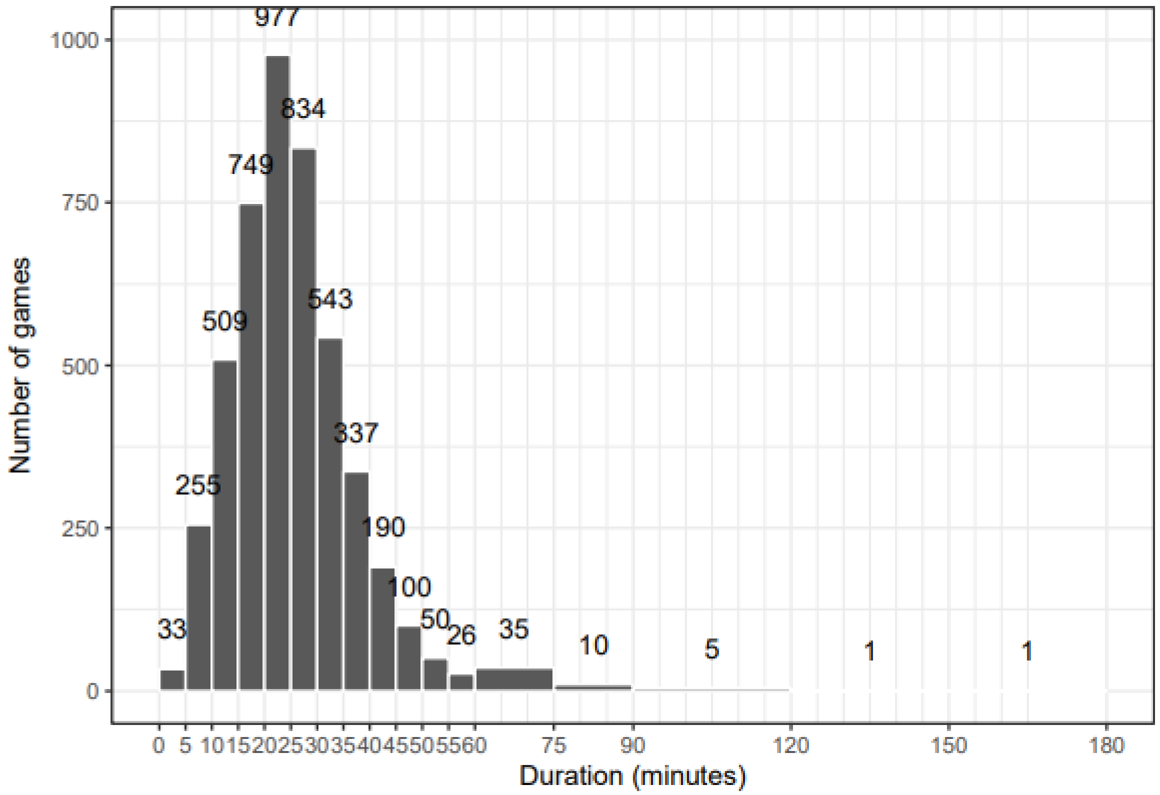


The Bias-Variance Trade-off II

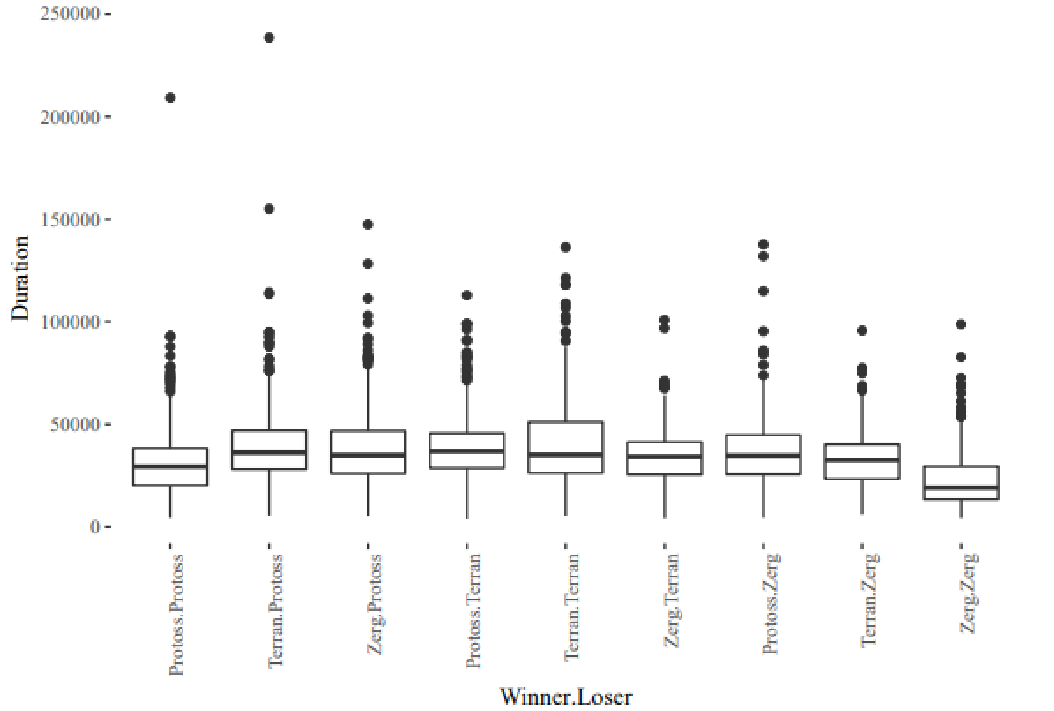
> **<u>Lesson 3:</u>** More flexible statistical methods (ex: regression spline) have more variance and less bias; more rigid statistical methods (ex: linear regression) have more bias but less variance.



Applications to StarCraft: Duration of Games



The mode is between 20 and 25 minutes, with average equal to 25.2570211 and median 25.2570211; they are not too different which indicates that the distribution is not too skewed.



And the chart shows that, except for the matches that include Zerg against Zerg and Protoss vs. Protoss, duration is quite similar and is more affected by the game dynamics that by differences or similarities between races.

Characteristics of Winners and Losers

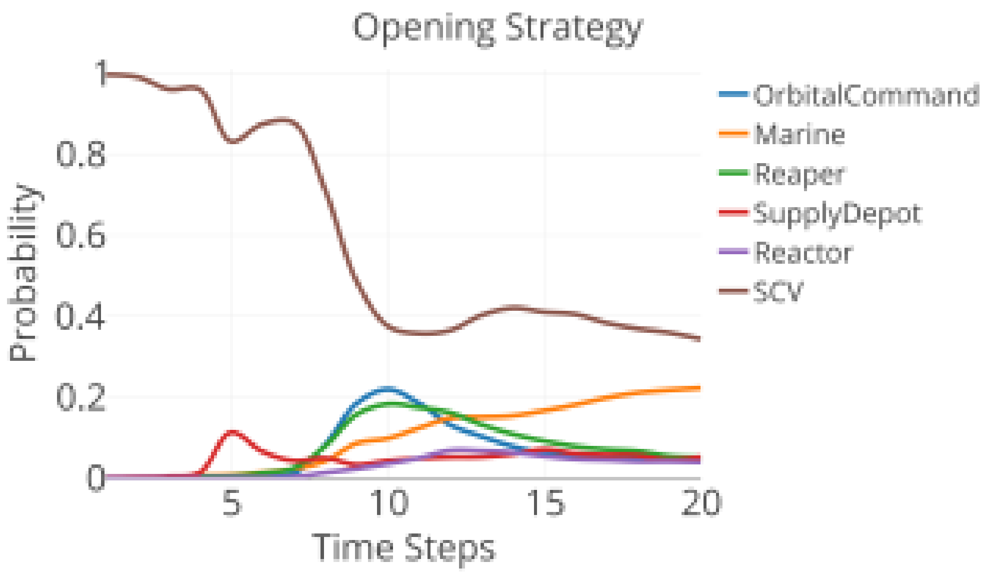


Figure 4: **Opening Strategy of the Winners.** The 6 lines show the probabilities of training a certain unit in the first 20 steps. Best viewed in color.

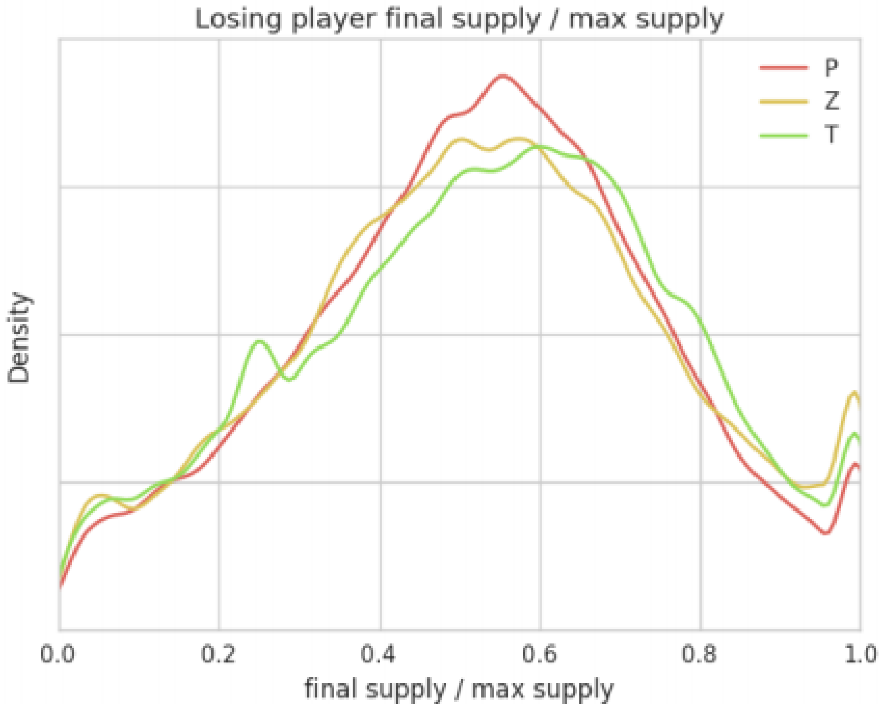
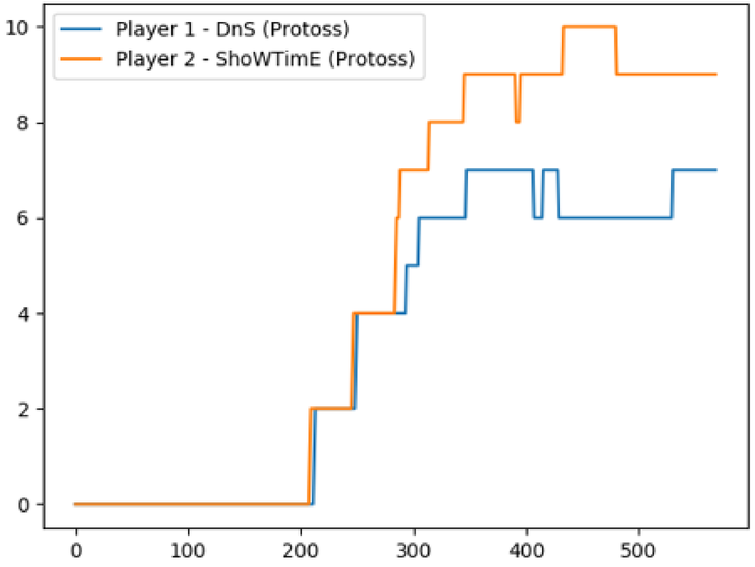
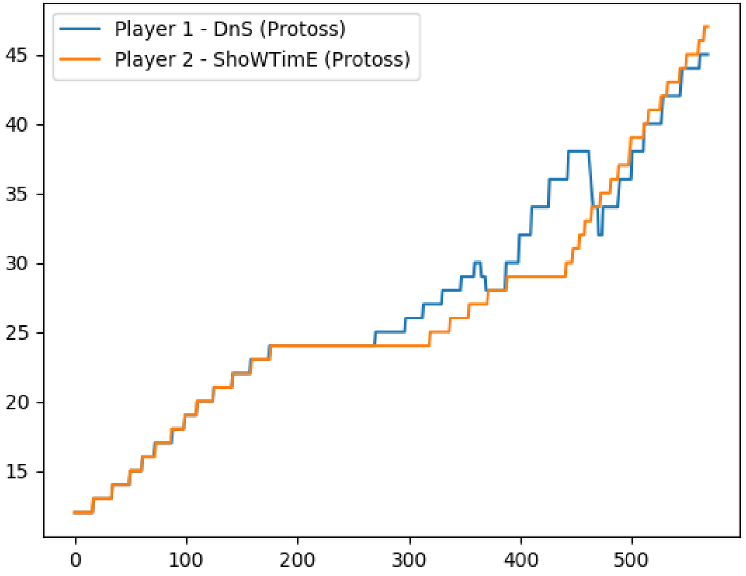


Figure 3: Ratio of effective supply to maximum supply of a surrendering player. A number near 0 indicates that the player fought until the bitter end. A number near 1 indicates that the player gave up fairly fast.

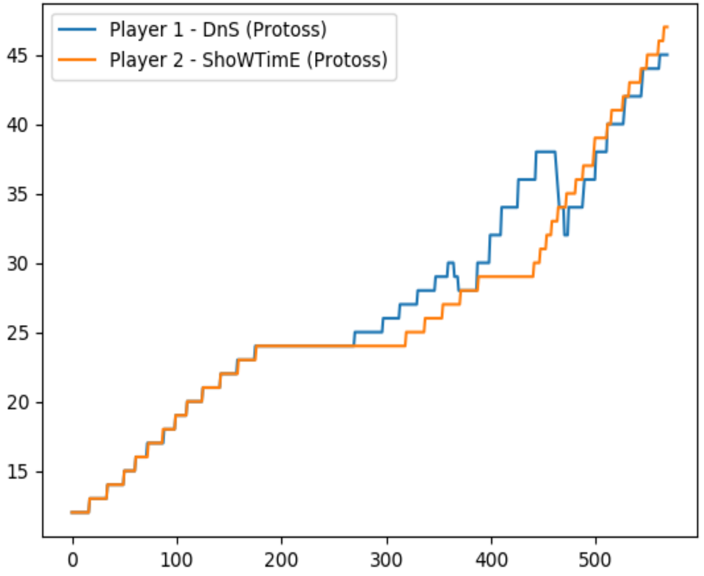
Plots from our StarCraft Data



Army Count



Workers Count



Cumulative Building Count