

# Howard Baek

(206)-591-4758 | [howardba@uw.edu](mailto:howardba@uw.edu)

[Blog](#) | [GitHub](#) | [LinkedIn](#)

## Education

### University of Washington

Master of Science in Biostatistics  
Pathway: Data Science

Seattle, WA  
March, 2023

### University of Washington

Bachelor of Science in Statistics  
Minor: Mathematics  
GPA: 3.70

Seattle, WA  
June, 2021

## Skills

- ◆ **Programming Languages:** R (proficient), Python (intermediate), SQL (intermediate)
- ◆ **Statistical Packages:** tidyverse (ggplot2, dplyr, tidyr, purrr, stringr), shiny, pandas, scikit-learn

## Work Experience

### National Oceanic and Atmospheric Administration

Machine Learning Intern

Remote

June 2021-Sept.2021

- ◆ Created an algorithm to automatically detect upwelling, an oceanographic process that has a significant impact on global fishery production.
- ◆ Applied principal component analysis, k-means clustering, and hierarchical clustering to real-world image data to group patterns of upwelling.
- ◆ Communicated findings in a [blog post](#) that elaborates on past work in upwelling, the detection algorithm, and the significance of the visualizations.

### Behavioral Research In Technology and Engineering (BRiTE) Center

NIH Research Assistant

Remote

Sept. 2020-June 2021

- ◆ Developed a **Shiny Dashboard** that allows patients and clinicians in addiction treatment to monitor patients' progress and goals over time.
- ◆ Implemented **Plotly** graphs that illustrate changes in patient-reported progress and goal measures over time.
- ◆ Programmed a **shinyMobile** application of the dashboard for optimal viewing on a smartphone.

### Upwork

Top-Rated Freelance Data Scientist

Remote

Oct. 2018-Aug.2019

- ◆ Consulted with an education startup to develop a learning system that measures student performance on quizzes.
- ◆ Created traffic flow **ggplot2** visualizations from sensor data that is useful for first responders in an emergency.
- ◆ Designed an internal facing dashboard that displays marketing metrics and a similar dashboard for a semiconductor company that is connected to a **MySQL** database.

### George Mason University Dept. of Computer Science

NSF REU Educational Data Mining Research Intern

Fairfax, VA

June 2018-Aug.2018

- ◆ Analyzed real world datasets from Stanford Lagunita's course, Statistics in Medicine, by creating exploratory visualizations in **ggplot2** and employing k-means clustering.
- ◆ Built an interactive web application (**shinydashboard**) that presents my analysis and visualizations in R.
- ◆ Presented a demo of the dashboard at the 9th International Learning Analytics and Knowledge (LAK) Conference.

### HM Health Solutions

Technology Intern

Pittsburgh, PA

May 2017-Aug. 2017

- ◆ Developed an internal Microsoft SharePoint homepage to redirect the Workforce Planning and Management department's customers to its services and a team page for the Talent Strategies team that supported recruiting and talent development.

## DataCamp

Content Partnerships Intern

Remote

Jan. 2018-Aug. 2018

- ◆ Researched 200 potential **Python** instructors by reviewing LinkedIn profiles, conference talks, and blogs.
- ◆ Worked with Copper, a CRM database, to maintain database of existing instructors or contacts within the company.

## Data Science Writing

---

### [Solutions to ggplot2: Elegant Graphics for Data Analysis](#)

- ◆ Produced a website hosting a solution manual to the exercises in the 3<sup>rd</sup> edition of ggplot2: Elegant Graphics for Data Analysis.

### [Animating Expected Possession Value in the NBA](#)

- ◆ Visualized each possession of a NBA game (using **gganimate**) within the context of Expected Possession Value (EPV), which is the number of points the offense is expected to score by the end of the possession in real time.

### [Quantifying Differences between the Regular Season and Playoffs using Survival Analysis](#)

- ◆ Employed survival analysis to find the difference in the rates of goals, shots, or hits from the regular season to the playoffs in hockey.

### [Hacking the NHL PBP App in Shiny](#)

- ◆ Walked through a month-long process building the National Hockey League Play-by-Play App with **shiny**, giving a behind-the-scenes look.

## Relevant Experience

---

### Honorable Mention at RStudio's 1<sup>st</sup> Shiny Contest

April 2019

- ◆ [First contest](#) at RStudio where **shiny** apps were judged based on technical merit and artistic achievement.

### Poster Presentation of Shiny Dashboard

June 2021

- ◆ [Poster](#) presented at the RSA/ISBRA-2021 Virtual Scientific Meeting, a conference for researchers in alcoholism and alcohol-related problems.

### Lightning Talk at CascadiaRConf 2021

June 2021

- ◆ [5-min talk](#) on developing a **Shiny Dashboard** and making it interactive for non-technical users.
- ◆ CascadiaRConf is an R conference serving the Pacific Northwest region (Oregon/Washington/BC).

### Invited Talk at Applied Analytics Club At UW

April 2020

- ◆ [1-hour online presentation](#) on the **gganimate** project explained in the blogpost, "Animating Expected Possession Value in the NBA".

## Leadership Experience

---

### Republic of Korea Army

Instructor of Reserve Forces

Seoul, South Korea

Aug. 2014-May 2016

- ◆ Received a recognition award for voluntary military service due to possession of Canadian citizenship.

## Extracurricular Activities

---

### UW Husky Ice Hockey Club

Statistician

Seattle, WA

Sept. 2019-June 2020

- ◆ Introduced a shot tracker application using **shiny**, a R package for interactive web applications, to record shot locations.