Trinity College Dublin

Biography of an Influential Software Engineer:

Dr. Hadley Wickham

By:

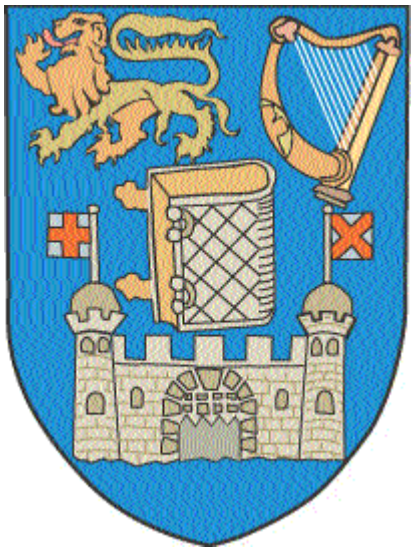
Amaya Aita

[15317530]

CS3012: Software Engineering

Prof. Stephen Barrett

Friday 10th November 2017



As I contemplated writing this essay, I realized early on that trying to decide on a single Software Engineer who influenced the world would be a challenge. On one hand, it is tempting to write about the young, glamorous and relevant, like Evan Spiegel. It would be easy to produce page upon page of impressive statistics and figures pertaining to his Snapchat empire. Yet, on the other hand, an overwhelming sense of gratitude almost persuaded me to write about the so-called fathers of Computer Science, without whom I wouldn't be studying the subject, much less have a module CS3012 to write a paper for. So, in the end I settled for a happy median- a young man from New Zealand, who sports an array facial piercings and brightly coloured shirts, yet, without him, one of the Top Ten most used languages worldwide would be at an unfathomable loss- Hadley Wickham.

Wickham is a software engineer, though those who know him say he's more likely to classify himself as a statistician. This is because of his work with the statistical programming language, R. Through his contributions, R has become an icon in data science circles. The core of the language has a reputation as a “powerful, yet occasionally counterintuitive language”, according to Wickham himself. He tackled, and continues to tackle, the counterintuitive parts of R by creating a number of the most used packages available for the language today. Wickham makes no secret of his love of the language, and in his path he has convinced thousands to follow him in his opinions. In the past, he has been quoted as saying “I strongly believe that R is the best language for data science, and I spend most of my time trying to make it even better” and “[many programmers] see R and think it is ridiculous and awful, but that didn’t happen to me... I thought it was really interesting”. The quantity *and* quality of Hadley's contributions to R are often discussed at data science events, and have turned him into a niche cult hero.

Wickham was born in New Zealand, and lived there throughout his youth. He expressed an interest in programming early on, and pursued this by securing a job at Microsoft as a teenager. In 1999, he went on to study the subject at University and, in 2004, he graduated with a Bachelor of Human Biology, B. Sc. (Statistics & Computer Science) and a M. Sc. (Statistics) from the The University of Auckland. In 2013, he revealed in an interview that he had originally planned to become a medial doctor, before quickly switching out of Biology, and into Statistics & Computer Science. The University of Auckland is not only his chosen place of study, but also the birth place of the R programming language, which was created in 1993 by statisticians Ross Ihaka and Robert Gentleman. However, it wasn't until Wickham left New Zealand to peruse a PhD at Iowa State University, that he started working on and developing packages for R. In 2006, in the midst of his PhD, he received the John Chambers Award for Statistical Computing. This was the first recognition his contribution to tools for data reshaping and visualization would receive. He completed his PhD in 2008, and moved to Rice University as an Assistant Professor. In 2013, Wickham joined Rstudio (a free and open-source integrated development environment for R) as a Chief

Scientist, where he currently works full time on contributions to the R ecosystem.

The Collection of packages that Wickham has developed are collectively known as the “tidyverse”. The name comes from the “tidy” approach to data operations which all of the tidyverse packages follow. This tidy principle, at its core, can be simply thought of as arranging data so that each column is a variable and each row is an observation in the statistical sense. Reshape was Wickham's first package, and was released as he was still a PhD student, yet it is still popular today. ggplot2 was then developed, by far his most popular package having been downloaded millions of times. It popularized a change in the way data visualizations are conceptualized. Charts created using ggplot2 are extremely prevalent in academic literature and reduction in bad academic graphical representations alone makes Wickhams work influential. The ggplot2 package is built on the theoretical work of Leland Wilkinson's Grammar of Graphics and allows for creation of graphics in a structured way that exposes features of the data in a natural way. There are many packages in the “tidyverse” including lubridate for dates, tidyr for tidying data, stringr for strings, httr for accessing web APIs, plyr for generalized function application readr for reading .csv and fwf files, readxl for reading .xls and .xlsx files, haven for SAS, SPSS, and Stata files, httr for talking to web APIs, rvest for scraping websites, xml2 for importing XML file, and much, much more.

The success of these open source packages, in particular ggplot2, is what lead to Wickham's move to full time software development at Rstudio. Since joining the enterprise, he has worked on many projects, one of which is a successor to ggplot2 called gvis. The integration of Wickham's ideas and packages, with other new technologies being developed at Rstudio, has pushed at the boundaries of what is possible with the R language. The package “Shiny”, developed at Rstudio, enables the creation of interactive web apps. This allows the work of Wickham to be viewed interactively by audiences that would have before now only been able to see the visual product of data analysis. Providing interactive gvis/ggplot2 charts through the web is an important step than allows analysis and reporting/journalism to exist together and not as separate stages.

Hadley Wickham is a young software engineer whose body of work continues to grow. In 5 or 10 year, I am confident his biography will be greatly ectended. In 2015, he was made a Fellow by the American Statistical Association for "pivotal contributions to statistical practice through innovative and pioneering research in statistical graphics and computing". He has been called “the Man Who Revolutionized R” and with more than double the downloads of the next closest contributor it would be hard to argue otherwise.

Some interviews with him have touched on his modest amount of fame and what it is like to be a “data celebrity”. He said, when asked about being recognized at conferences and events, “I can see my level of fame getting to the point where it’s actively uncomfortable” and wishes someone could write about “how to be a celebrity in a very specific field”, but goes on

to say he doesn't need to worried about being recognized on the street, with the exception of when in sunny San Francisco.

Recognizing the work of people who are exceptionally successful in life is not only useful, but it focuses effort. When the achievement that we idolize is too great, we are often looking a the winner of a lottery. People like Hadley Wickham show that through hard work it is possible to change a corner of the world.

References and Sources:

1. *http://hadley.nz/*

2. *http://had.co.nz/portfolio/cv.pdf*

3. *https://www.linkedin.com/in/hadleywickham/*

4. *https://www.quora.com/How-is-Hadley-Wickham-able-to-contribute-so-much-to-R-particularly-in-the-form-of-packages*

5. *https://priceonomics.com/hadley-wickham-the-man-who-revolutionized-r/*

6. *http://www.huffingtonpost.ca/murtaza-haider/data-visualization-media_b_7339342.html*

7. *https://www.biostars.org/p/115481/*

8. *https://statr.me/2013/09/a-conversation-with-hadley-wickham/*

9. *https://qz.com/1007328/all-hail-ggplot2-the-code-powering-all-those-excellent-charts-is-10-years-old/*