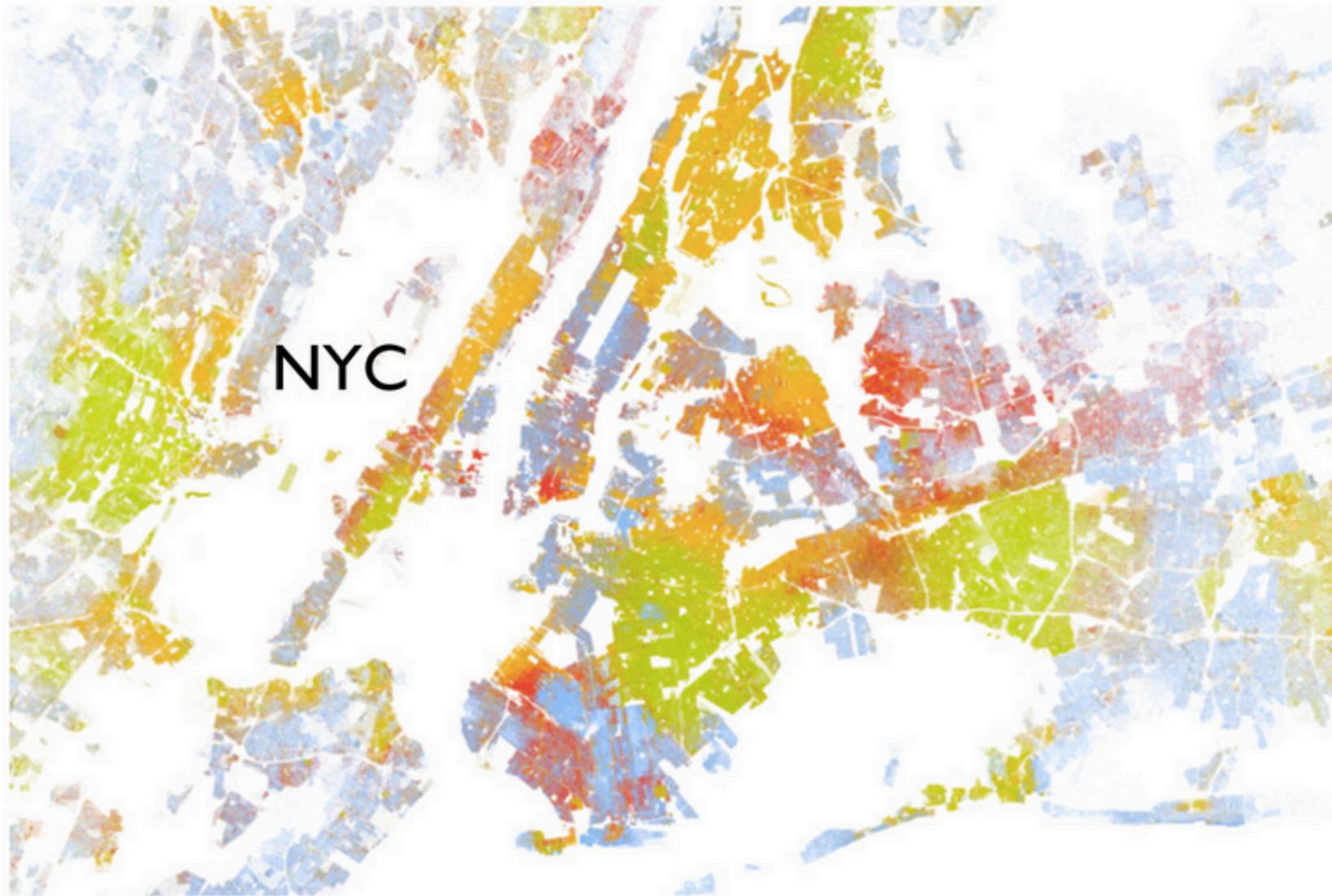


#dataviz movement is toward transparency

This is awesome for us
because many, many
resources are available
online

The Best Map Ever Made of America's Racial Segregation

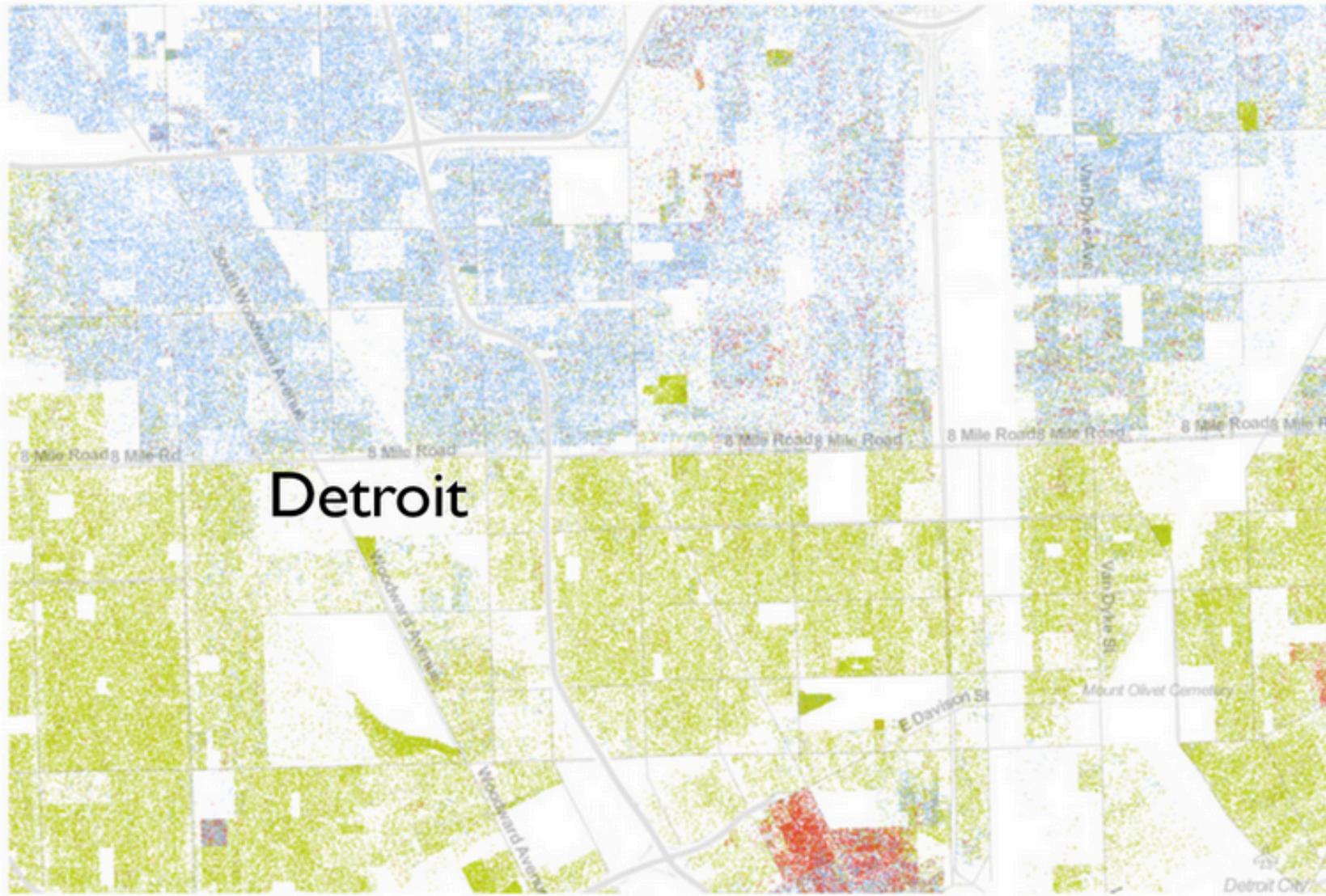
<http://www.wired.com/design/2013/08/how-segregated-is-your-city-this-eye-opening-map-shows-you/?viewall=true>



This map, created by Dustin Cable at University of Virginia's Weldon Cooper Center for Public Service, is [the most comprehensive representation of racial distribution in America ever made](#). Here: New York City. *Image: Dustin Cable* White: blue dots; African American: green dots; Asian: red; Latino: orange; all others: brown

The Best Map Ever Made of America's Racial Segregation

<http://www.wired.com/design/2013/08/how-segregated-is-your-city-this-eye-opening-map-shows-you/?viewall=true>



In Detroit, among the most segregated cities in America, 8 Mile Road serves as a sharp dividing line. *Image: Dustin Cable White: blue dots; African American: green dots; Asian: red; Latino: orange; all others: brown*

This is the most comprehensive map of race in America ever created.

White people are shown with blue dots; African-Americans with green; Asians with red; and Latinos with orange, with all other race categories from the Census represented by brown. Since the dots are smaller than pixels at most zoom levels, Cable assigned shades of color based on the multiple dots therein. From a distance, for example, certain neighborhoods will look purple, but zooming-in reveals a finer-grained breakdown of red and blue—or, really, black and white.

"There are a lot of moving parts in this process, so this can cause different shades of color to appear at different zoom levels in really dense areas, like you see in NYC," Cable explains. "I played around with dot size and transparency for a while and settled on the current scheme as being adequate." You can [read more about Cable's methodology here](#), but it comes down to this: When you're dealing with 300 million dots at varying levels of zoom, getting the presentation just right is as much an art as a science.



The Racial Dot Map

One Dot Per Person for the Entire United States

Created by Dustin Cable, July 2013



Cool result is accompanied by explanation of how it was done

Methodology

Python was used to read the 50 state shapefiles (with the merged SF1 data). The GDAL and Shapely libraries were used to read the data and create the point objects. [The code](#) retrieves the population data for each census block, creates the appropriate number of geographic points randomly distributed within each census block, and outputs the point information to a database file. The resulting file has x-y coordinates for each point, a quadkey reference to the Google Maps tile system, and a categorical variable for race. The final database file has 308,745,538 observations and is about 21 GB in size. The processing time was about five hours for the entire nation.

The database file was then sorted by quadkey and converted to a .csv format. SAS was able to do this within an hour without crashing.

Processing 2.0.1 for 64-bit Windows was used to create the map tiles. The [Java code](#) reads each point from the .csv file and plots a dot on a 512x512 .png map tile using the quadkey reference and x-y coordinates. The racial categorical variable is used to color-code each plotted dot. This process used the default JAVA2D renderer, but other platforms may work better using P2D. Map tiles were created for Google Maps' zoom levels 4 through 13 to make the final map. A non-color-coded map was also produced to help add more contrast for lightly populated areas. In total, the color-coded and non-color-coded maps contain 1.2 million .png files totaling about 7 GB. Producing all of the map tiles in Processing took about 16 hours for the two maps.

The Google Maps API is used to display the map tiles. Map tiles with zero population are never created using the above method. Therefore, [an index was used](#) to tell the map application whether a tile exists in order to prevent 404 errors.

The entire code is up on [GitHub](#) and was adapted from code developed by [Brandon Martin-Anderson](#) and [Peter Richardson](#) in order to account for the racial coding and errors in reading the shapefiles.

#dataviz movement is toward reproducibility

This is awesome for science and engineering, because we don't need another source of confusion

HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE
EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?

Automation can save time

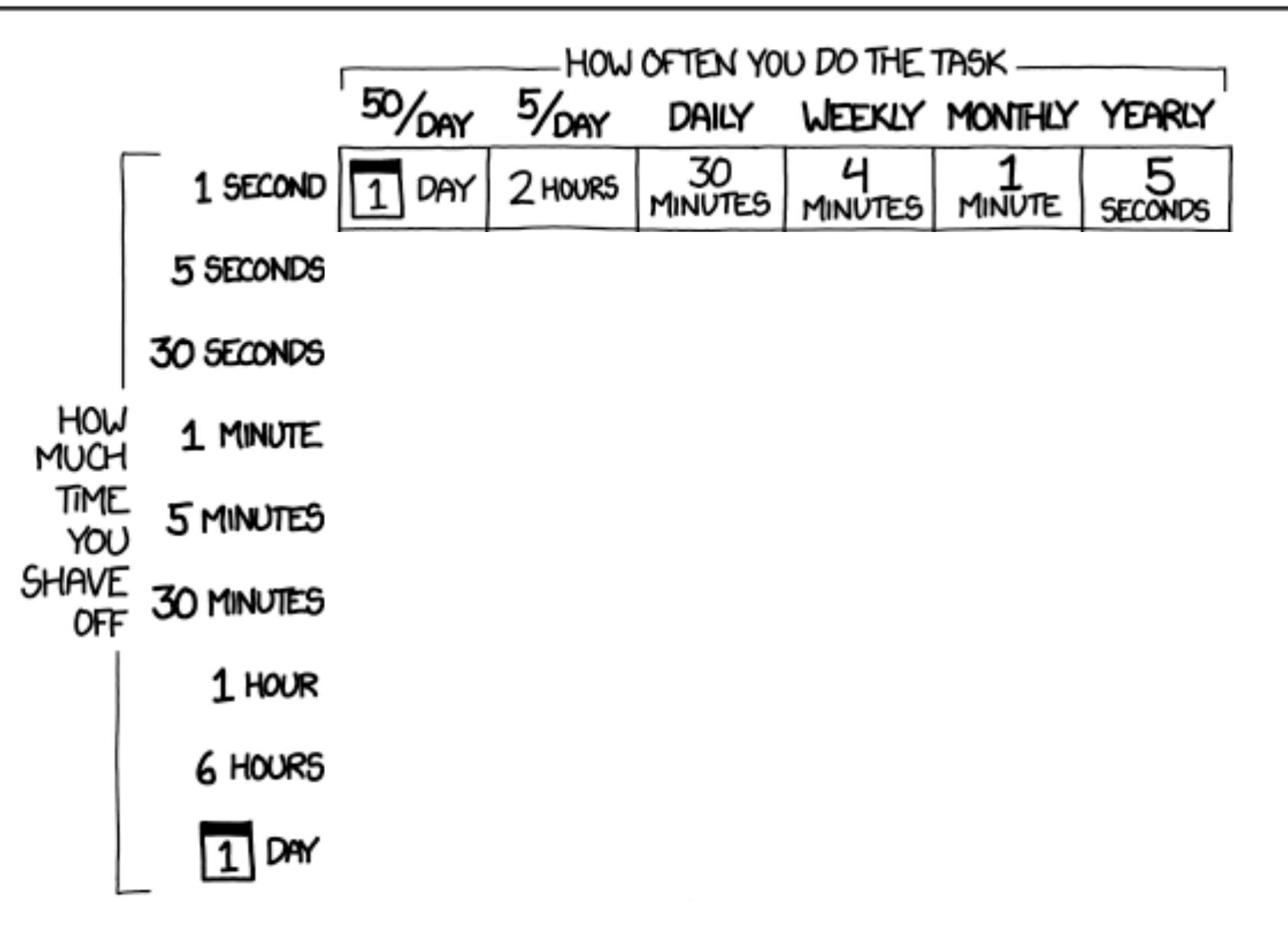
xkcd 'Worth the time',
<https://xkcd.com/1205/>



the university
for the real world

Automation can save time

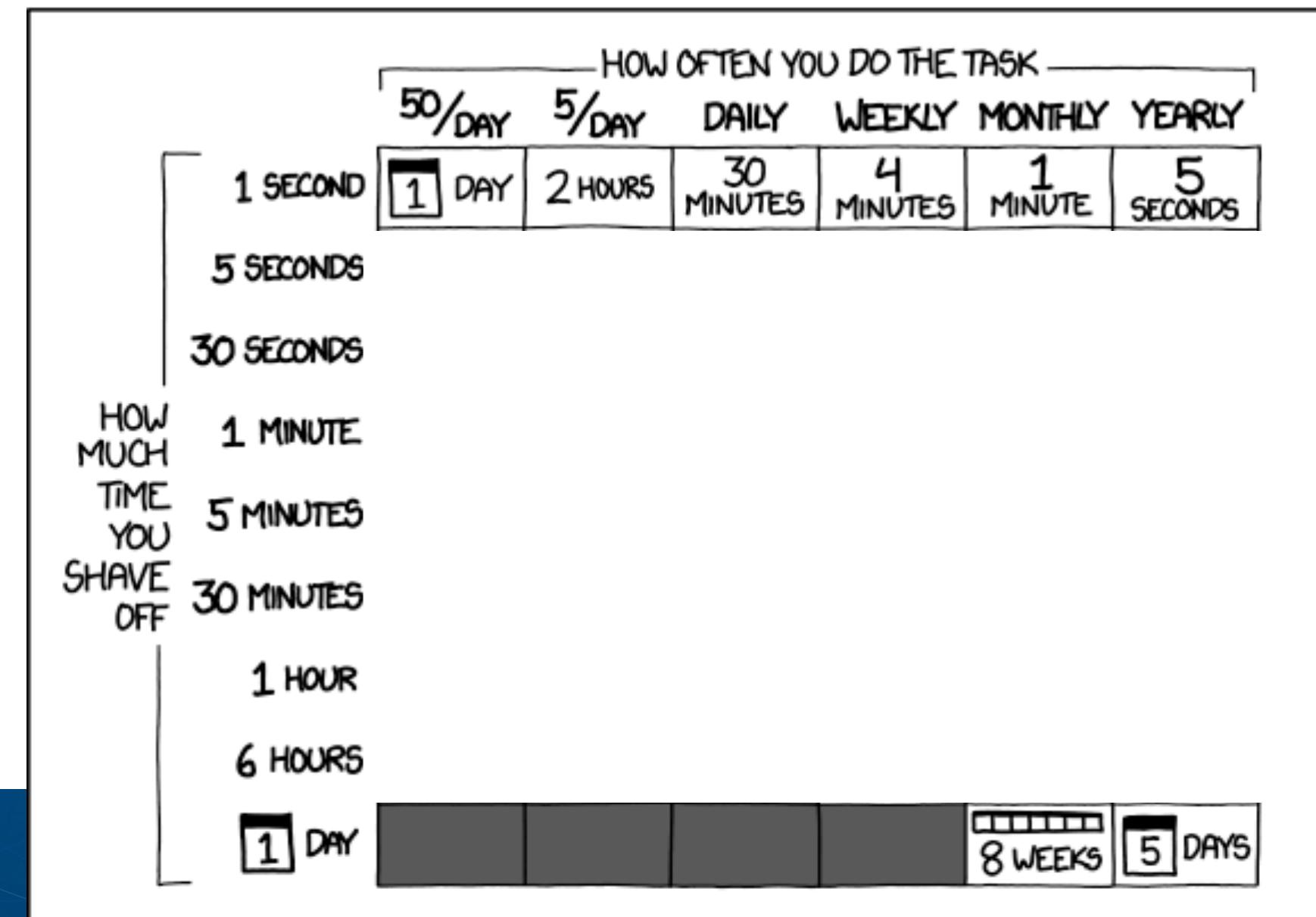
HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE
EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?
(ACROSS FIVE YEARS)



Automation can save time

xkcd ‘Worth the time’, <https://xkcd.com/1205/>

HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE? (ACROSS FIVE YEARS)



Automation can save time

xkcd 'Worth the time',
<https://xkcd.com/1205/>
QUT the university
for the real world

HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE
EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?
(ACROSS FIVE YEARS)

		HOW OFTEN YOU DO THE TASK					
		50/DAY	5/DAY	DAILY	WEEKLY	MONTHLY	YEARLY
HOW MUCH TIME YOU SHAVE OFF	1 SECOND	1 DAY	2 HOURS	30 MINUTES	4 MINUTES	1 MINUTE	5 SECONDS
	5 SECONDS	5 DAYS	12 HOURS	2 HOURS	21 MINUTES	5 MINUTES	25 SECONDS
	30 SECONDS	4 WEEKS	3 DAYS	12 HOURS	2 HOURS	30 MINUTES	2 MINUTES
	1 MINUTE	8 WEEKS	6 DAYS	1 DAY	4 HOURS	1 HOUR	5 MINUTES
	5 MINUTES	9 MONTHS	4 WEEKS	6 DAYS	21 HOURS	5 HOURS	25 MINUTES
	30 MINUTES		6 MONTHS	5 WEEKS	5 DAYS	1 DAY	2 HOURS
	1 HOUR		10 MONTHS	2 MONTHS	10 DAYS	2 DAYS	5 HOURS
	6 HOURS				2 MONTHS	2 WEEKS	1 DAY
	1 DAY					8 WEEKS	5 DAYS

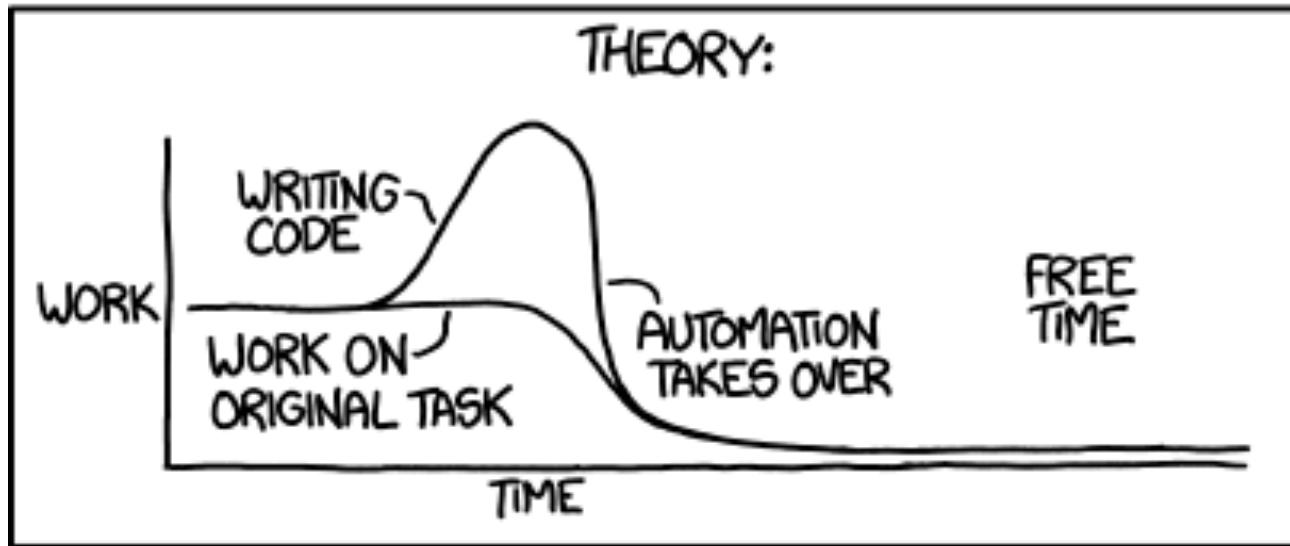
"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"

Automation can waste time

xkcd 'Automation',
<https://xkcd.com/1319/>
 the university
for the real world

Automation can waste time

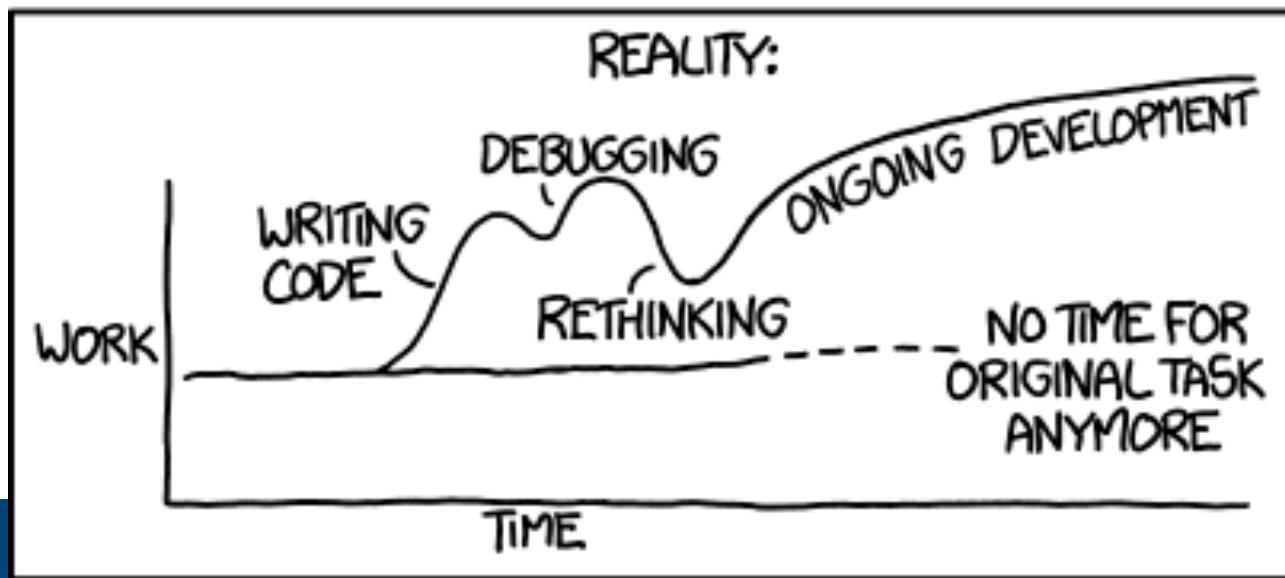
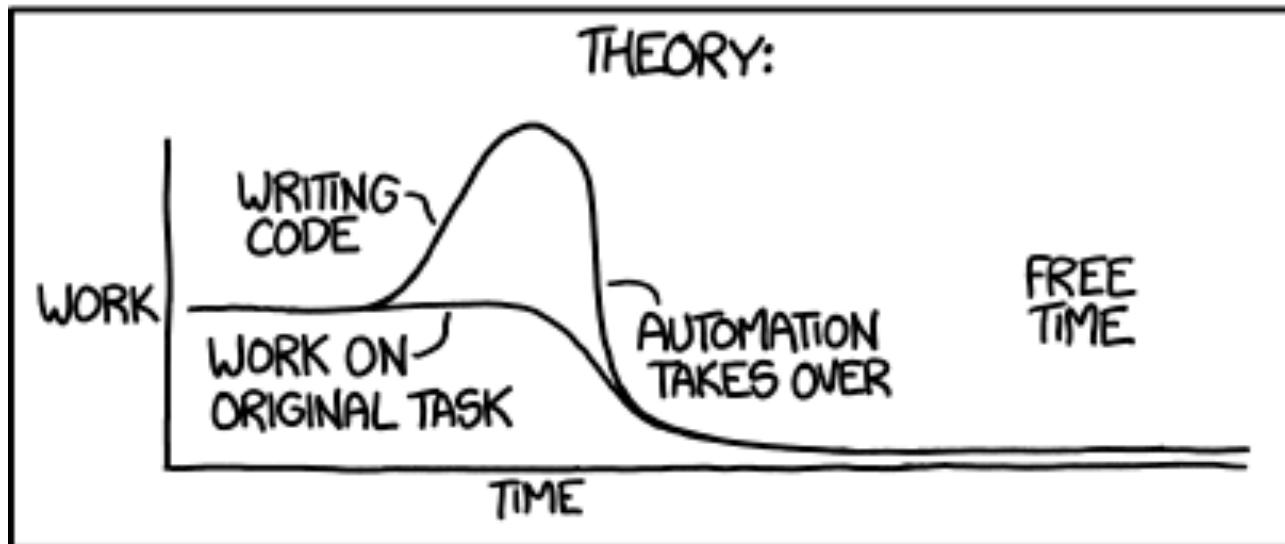
"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



xkcd 'Automation',
<https://xkcd.com/1319/>

Automation can waste time

"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



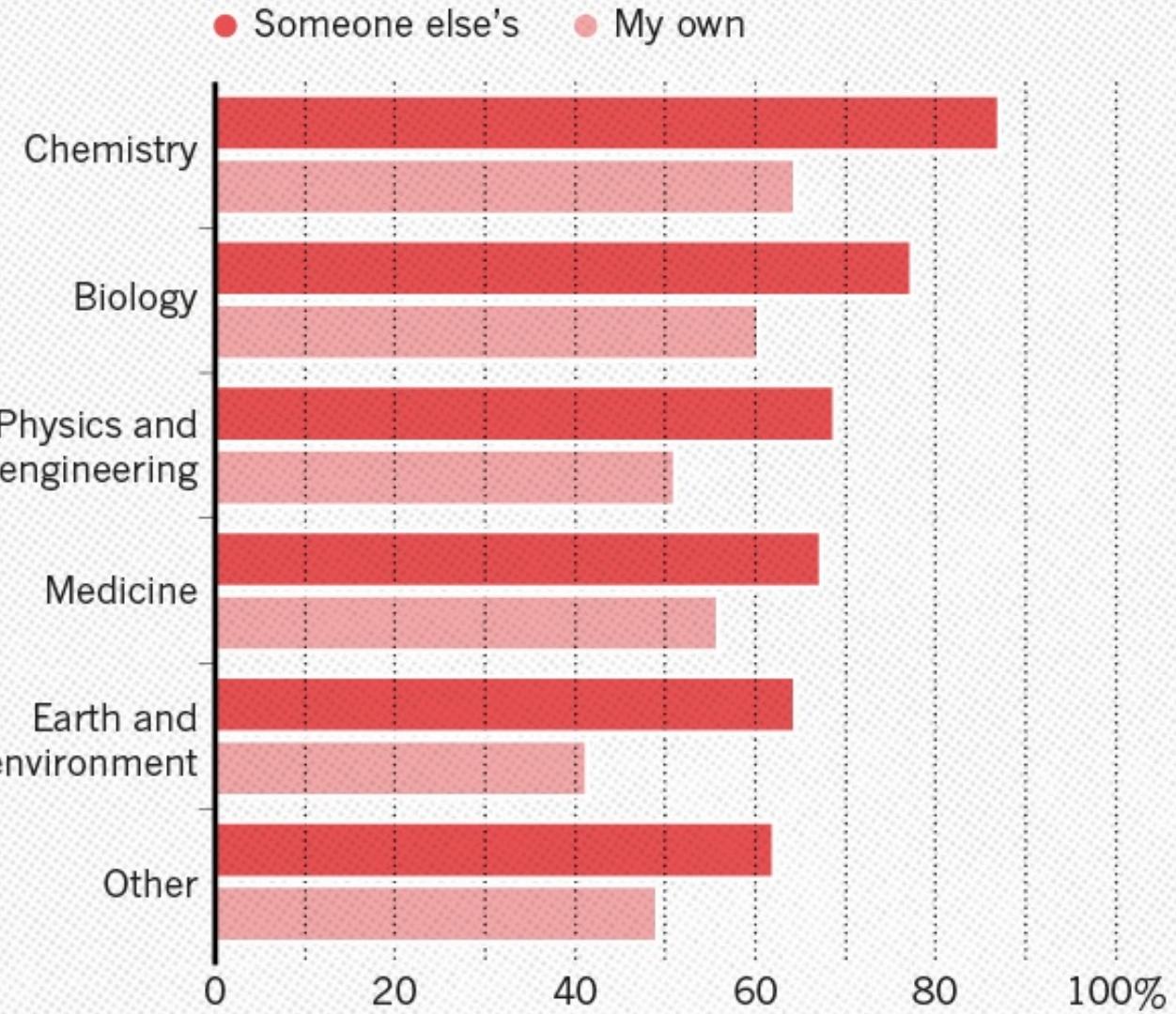
xkcd 'Automation',
<https://xkcd.com/1319/>

Automation = Reproducibility

Baker (2016) 1,500
scientists lift the lid on
QUT reproducibility. Nature
the university for the real world

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



This unit:

Real data, real tools

- Things will break
- Things will change

Imposter Syndrome

