

Turning challenges into opportunities as an isolated data scientist

Lessons from rstudio::conf

Monica Gerber, MPH

2018-05-08

This talk

1. Isolated data scientists¹
2. Jumping into the R community (w/ links to rstudio::conf)
3. My shiny app

How can isolated data scientists at Partners address the challenges and take advantages of the opportunities of these roles?

[1] Baumer, B.S., 2017. Lessons from between the white lines for isolated data scientists. [PeerJ PrePrints](#).

My background

I am currently a Biostatistician at:

- CCHI



- General Academic Pediatrics



- Anthropologist turned public health data scientist
- Interests: R, statistical methods for causal inference, community health, social determinants of health
- R experience:
 - Once in undergrad stats course
 - Started learning in 2015 w/ [swirl](#) package and [coursera](#)

The landscape

Many "isolated data scientists" at Partners.

My analysis:

- Many opportunities
- Great challenges may overwhelm these opportunities



Problem

- Training is focused on theoretical topics (how to do a t-test, etc.)
- Not enough focus on the *practice* of being a data scientist
- When you're on your own, logistical challenges can be really tough



Options

1. Remain in isolation. Use SAS.
2. Jump into R community and focus on the *practice* of being a data scientist.
 - Build what you need
 - Organize your workflow
 - Find joy in what you do
 - Be productive while learning new things



rstudio::conf

SAN DIEGO

Build what you need

No need to wait

Just get your idea out there

Talks to watch

1. Contributing to tidyverse packages
2. Fireside chat with Rstudio
3. You can make a package in 20 minutes
4. infer: a package for tidy statistical inference

Organize your workflow

- Tidyverse packages
- Learn iteration



Talks to watch

1. The lesser known stars of the tidyverse
2. Modeling in the Tidyverse
3. Data rectangling
4. Tidy eval: Programming with dplyr, tidyr, and ggplot2
5. Teach the Tidyverse to beginners

Find joy in what you do

- Reach out
- Make things fun



Talks to watch

1. A SAS-to-R success story
2. Five packages in five weeks – from boredom to contribution via blogging
3. Imagine Boston 2030: Using R-Shiny to keep ourselves accountable and empower the public

Balance productivity and learning

- A lot to learn from outside your discipline
- With Shiny, we're stepping more into a new type of role



Talks to watch

1. Rapid prototyping data products using Shiny
2. Agile data science
3. Many other Shiny talks



getting to work with Shiny

My first app

- An app to visualize economic hardship in Massachusetts based on methodology from Brookings Institution's analysis of **Chicago** neighborhoods.
- Center for Community Health Improvement works in low income neighborhoods where economic instability contributes to poor health outcomes.

My top logistical challenges¹

Marginality

- Data science is tangential to the core work.
- Community health workers and managers are most concerned about addressing the immediate problems of patients. Leaders are often focused on finding support for the the community health work.

Mystery

- People aren't sure exactly what it is that data scientists do.
- What is Shiny? What is R? Can you help me with my spreadsheet?

Misalignment

- Poor placement of data scientists.
- At MGH, people are working many different health centers and offices.

[1] From Elaine McVey's talk and blog post: <http://www.datawoman.com/>

Marginality -> user stories

User stories are a way to describing the work that puts the focus on the value to the end user.

Before:

I am going to recalculate the Chicago hardship index using 2015 census data to show how economic hardship varies across Massachusetts and visualize it with a choropleth map.

Marginality -> user stories

User stories are a way to describing the work that puts the focus on the value to the end user.

Before:

I am going to recalculate the Chicago hardship index using 2015 census data to show how economic hardship varies across Massachusetts and visualize it with a choropleth map.

As a ____, I need to ____, so that I can ____.

After: As an executive at MGH, I need to build a case for CCHI, so that I can find donors to support our work.

Mystery -> vertical slicing

Maximize the work not done.

Find the thinnest vertical slice.

- Instead of spending hours and hours on something, and then presenting this to the user (the whole cake), the idea is to find the thinnest vertical slice.
 - Quickly get from the thinnest layer to the top
 - Narrow the scope
- Make use of all the time you spend on data collection and manipulation.



Ideas for vertical slicing with shiny

- Build 1 page at a time
- Use flexdashboards
- Don't worry if it's not pretty
- Don't worry if you don't do everything the "right" way



Misalignment -> Stakeholder reviews

- Review the vertical slices.
- What do you want to spend time on vs. what is creating value?



My app

[Link](#) to my app.

Economic Hardship in Massachusetts

What is the hardship index?

The Hardship Index is a measure of economic hardship. It utilizes six indicators to provide a more comprehensive view of economic hardship than single indicators. The calculation is based on the "Intercity Hardship Index," by Richard P. Nathan and Charles F. Adams, Jr. in "Understanding Urban Hardship," Political Science Quarterly 91 (Spring 1976): 47-62.

The economic hardship score is the median of six variables that have been standardized on a scale of 0 to 100. The six variables include:

1. Unemployment (over the age of 16 years),
2. Education (over 25 years of age without a high school diploma)
3. Per capita income level
4. Poverty (below the federal poverty level)
5. Crowded housing (housing units with more than one person per room)
6. Dependency (population under 18 or over 64 years of age)

Higher hardship index scores indicate worse economic conditions.

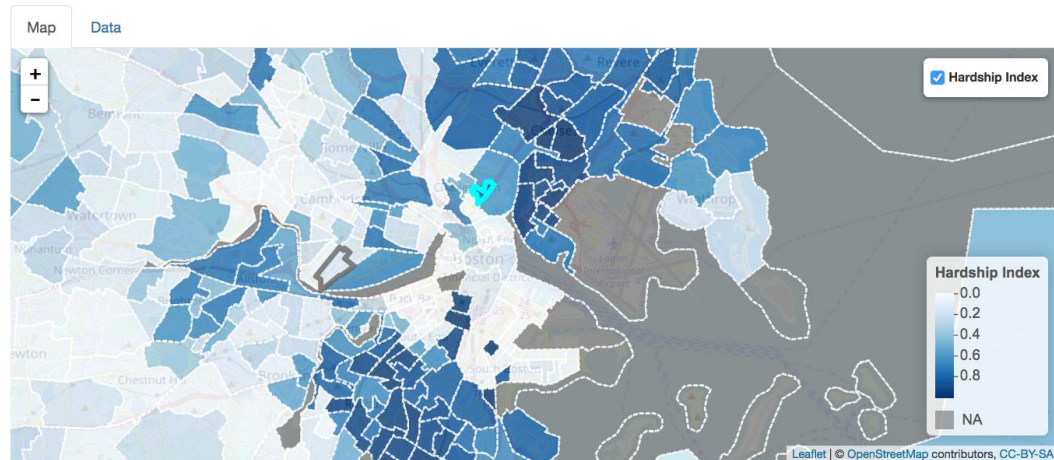
The hardship index in MA

The Economic Hardship Index for MA utilizes the American Community Survey 2015 5-year estimates. It includes indicators for Census Tract Blocks in Massachusetts. Census tracts with small populations (under 500) were excluded from the analysis.

More information

This Shiny app is still under development. Monica Gerber conducted the analysis and developed the app.

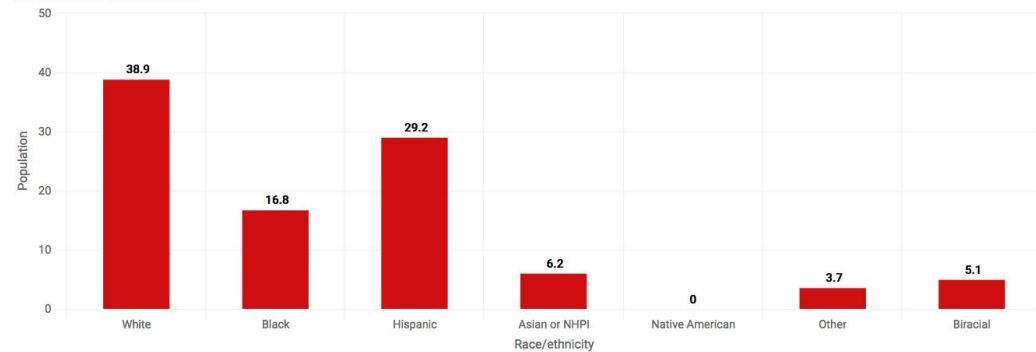
Please contact Monica Gerber with any questions



Click a census tract for more information.

Census Tract 402, Suffolk County, Massachusetts

Hardship Index: 0.91, Population: 1841



Key takeaways

- R is a great way to address the challenges of isolation.
- Use it because it's fun.
- Tons of online and offline communities to support you.
- Learning is hard. Be kind to yourself & have compassion for others.



Thank you & keep in touch.

mwgerber@mgh.harvard.edu

Twitter: @mwgerber

Github: @monicagerber

Partners Gitlab: @mwg11

MGH Center for Community Health Improvement

MGHfC General Academic Pediatrics

R-Ladies Boston*

*Next meetup on May 30th w/ Mara Averick.