

Proposal to Create an R Consortium Working Group Focused on US Census Data

The Problem

R users who wish to work with US Census Data face two significant problems: Package Selection and Data Navigation.

Package Selection

An R user looking for packages to help with a project on US Census Data would likely start by going to CRAN's [list of contributed packages](#) and doing a search for "Census". This process is non-optimal for a number of reasons:

1. It yields 41 hits, which is a large number to sort through.
2. Many package titles appear to be duplicative (e.g. `censusr`'s title "Collect Data from the Census API" is very similar to `censusapi`'s title of "Retrieve Data from the Census APIs").
3. Some packages that deal with US Census Data do not have the word "Census" in their name or title (e.g. the `choroplethr` package).

For these reasons, even an experienced R user might find it challenging to determine which R package, if any, can help them with their Census-related project.

Note that these issues might also lead to package duplication on CRAN. That is, a developer might create a package to solve a problem which an existing package already solves.

Data Navigation

People often turn to the Census Bureau for answers to questions such as "How many people live in a certain region?" They are often surprised that the answer depends on which data product they query. That is, many R users begin their interaction with Census Data not fully understanding the array of data available to them, or when it is better to use one data product over another.

The Plan

We would like to create an R Consortium Working Group to foster greater cooperation between the US Census Bureau and R community. We hope that the working group will become a place where statisticians working with census data can cooperate under the guidance of the Census Bureau.

The first project we would like the Working Group to take on is the creation of a guide to getting started with Census data in R ("The Guide"). The Guide would address the problems listed above by:

1. Listing, describing and linking to packages currently available via CRAN that work with US Census Data.
2. Listing, describing and linking to reference materials for better understanding and working with Census data.

The most likely failure mode for the **Package Selection** section is not including all the relevant packages, or not communicating the differences between the packages in a way that helps users decide which package is most appropriate for their use. At this point we do not know whether the Guide should simply copy the CRAN package description, or also include additional information. (And if it should include additional information, what information should that be?) We plan to address this risk by publishing and publicizing drafts of our work online, and incorporating feedback from the R community.

The most likely failure mode for the **Data Navigation** section is not providing resources which are useful or relevant to the actual needs of the R community. In the same way that CRAN has a wealth of packages that can be difficult to navigate, the Census Bureau also has a wealth of training material that can be difficult to navigate. We plan to address this by publishing and publicizing drafts of our work online, and incorporating feedback from the R community.

Another failure mode which we do not address in this proposal is maintenance of the Guide. While the Guide might be completely accurate the time of publication, it will naturally become less accurate over time. At this point it is not clear what the best way to maintain the Guide is.

The Team

Ari Lamstein. Ari is an R Trainer and Consultant who is currently working as a Data Science Instructor at the US Census Bureau. He has written several R packages that deal with US Census Data. Ari is planning to focus on on the **Package Selection** portion of the Guide.

Logan T Powell. Logan is the Developer Experience and Engagement Lead at the US Census Bureau. Logan is planning to work on the **Data Navigation** portion of the Guide.

Zach Whitman, PhD. Zach is the Chief Data Officer at the US Census Bureau. Zach is planning to work on future projects related to changes in the Census API.

Kyle Walker, PhD. Kyle is Associate Professor of Geography at Texas Christian University. Kyle is the primary author of the `tigris`, `tidycensus`, and `idbr` R packages for working with Census Bureau spatial and tabular data in R.

Project Milestones

Milestone 1: Select Publication Technology (1 Month, \$500)

Our first task will be selecting technology to use to publish the Guide. We would like the technology to be free and easy to access, as well as free to host and easy to update. We are planning to start this phase by evaluating Github, Github Pages and WordPress.

This milestone will be completed when there is a "Hello World" version of the Guide published, and both authors understand the workflow for editing it.

Milestone 2: Assemble list of resources to include (1 Month, \$1,000)

The initial package and resource lists will be based on our personal knowledge and experience. The lists will be stored in github so that other people can contribute to the lists. That github repository will then be publicized via our blogs and social media accounts.

This milestone will be completed when we have the final list of Package and Training resources that we plan to include in the guide.

Milestone 3: Complete Publication of Guide (1 Month, \$2,000)

After we have the final list of resources we plan to include in the Guide, we will need to write up the results.

This milestone will be completed once the final draft of the Guide is published.

We plan to announce completion of this milestone on our blogs and social media accounts.

Milestone 4: Complete Dissemination (1 Month, \$500)

Once the Guide is completed, we will focus on disseminating it to the largest possible audience.

We will start by simply announcing the Guide's completion on our blogs and social media accounts.

We will also reach out to the Census Bureau, which has already indicated an interest in linking to it from their website.

We also believe that CRAN might want to link to the Guide on the [Official Statistics & Survey Methodlogy Task View](#). However, we have not yet spoken to the Task View Maintainer about this. If the maintainer thinks that this project is not a perfect fit, then we are open to creating a separate task view dedicated to US Census Statistics.

How Can The ISC Help

We are seeking to create an ISC Working Group to promote using R to analyze US Census Data. The individual who will be committed to leading and managing the group's activities is Ari Lamstein.

We are requesting a \$4,000 grant to fund completion of the Working Group's first project: a Guide to getting started with Census Data in R ("The Guide"). The *Project Milestones* sections contains a breakdown of funding by milestone.

The Census Bureau is currently planning changes to its API. We hope that a future project for the Working Group will be to get feedback on the API from the R Community.

In addition to creating a Working Group and financially supporting the creation of the Guide, we believe that the R Consortium can help this project to succeed by facilitating cooperation among stakeholders and disseminating and promoting the Guide on the R Consortium website.

Dissemination

We plan to publish the Guide under the [Creative Commons Attribution 4.0 License](#).

As indicated in the *Project Milestones* Section, we plan to include the community in the drafting of the Guide, and we plan to publicize project milestones on our blogs and social media accounts.

The Census Bureau has also indicated an interest in linking to the Guide on its website once it is completed.

We also hope to speak with the maintainer of CRAN's [Official Statistics & Survey Methodology Task View](#) about including a link to the Guide. If the maintainer thinks that this project is not a perfect fit, then we are open to creating a separate task view dedicated to US Census Statistics.