

# StatPREP: Instructor Notes

*Daniel Kaplan*

*2018-06-04*



# Contents

<b>1</b>	<b>Pocket guide to StatPREP commands</b>	<b>5</b>
<b>2</b>	<b>Signing up for cloud services</b>	<b>7</b>
2.1	Google . . . . .	7
2.2	GitHub . . . . .	7
2.3	RStudio cloud . . . . .	8
<b>3</b>	<b>StatPrep Annie</b>	<b>9</b>
<b>4</b>	<b>Your course web site</b>	<b>11</b>
4.1	Setting up your web site . . . . .	11
4.2	Customizing the site content. . . . .	12
<b>5</b>	<b>Class data using Google Sheets</b>	<b>15</b>
5.1	Resources . . . . .	15
5.2	Setting up a new spreadsheet . . . . .	16
<b>6</b>	<b>Using RStudio.cloud</b>	<b>19</b>



## Chapter 1

# Pocket guide to StatPREP commands

**goal** ( **formula** , data = **table** , **options** )

```

counts(      ~ cat1
props(      ~ cat1
props( cat1 ~ cat2
df_stats(   ~ x
df_stats( x ~ cat1
gf_point(   y ~ x
gf_jitter(  y ~ x
gf_smooth(  y ~ x
gf_lm(      y ~ x
gf_density( ~ y
gf_histogram( ~ x
gf_boxplot(x~ cat1
gf_violin(x ~ cat1
gf_col(     x ~ cat1
facet: x ~ y | cat1

```

x	y	cat1	cat2
3.6	6	A	red
1.2	9	A	blue
12.0	2	B	blue

**Quick facts on data frames**  
**names(D) nrow(D) head(D)**

*Functions for randomization*  
**Trials←do(100)\*{task}**  
 Confidence intervals  
**data=resample(df)**  
 Permutation test  
**x ~ shuffle(cat1)**

*optional arguments*  
 graphics  
**fill = "blue"**  
**fill = ~ cat1**  
**color = ~ cat1**  
**alpha = 0.5**  
**jitter**  
**width = 0.2**

for df\_stats  
 mean ci.mean  
 median ci.median  
 sd ci.sd  
 coverage  
 min Q1 Q3 max  
 length

## Pocket guide to StatPREP R

rev 5-25-18

Figure 1.1:

## Chapter 2

# Signing up for cloud services

You don't need to install any special software on your own computer. Instead, we'll use services in the cloud that work through a standard web browser.

It's helpful if you set up a personal account on the cloud services listed below. That will save time on the day of the workshop and you can be a resource for your neighbor if they haven't had a chance to do so. (Do remember to keep track of your user ID and password. Writing it down is a good idea; you can change the password after the workshop if you are worried about security.)

### 2.1 Google

You may already have a Google account: many people have an account already or work at an institution that provides email and other services through Google. If so, you are all set.

If you don't already have an account, follow this link to sign up.

Setting up a Google account is entirely to streamline authentication to other services that we use with StatPREP. You do not need to change anything about your existing email service or how you use it.

### 2.2 GitHub

Funny name, huh? GitHub is a free service with tens of millions of users. It's most closely associated with software development, but our main use for it will be to give you a way easily to create a web page to give your classes access to whichever StatPREP tutorials, lessons, or Little Apps you choose to use with them.

Your institution may already provide you with a web site or a system such as Moodle or Blackboard that gives you a class-specific web page. If so, the point of setting up a GitHub account to use at the workshop is to make it easier for us to avoid having to figure out how to upload documents to a multiplicity of different web platforms.

In selecting your user ID for GitHub, keep in mind that the ID is something that will be visible to students. So, **ProfJones** or something of that kind is probably better than **Red\_hot\_pepper\_dude**.

Follow this link to GitHub's account creation page. And *don't be intimidated* by the "Built for developers" label.

## 2.3 RStudio cloud

We want you to have access to RStudio so that you can use it *if you decide you want to*. We'll show everyone some basics at the workshop so you can make an informed decision.

You can sign in using either your Google or your GitHub credentials; there's no need to set up a separate ID or password. Go to `rstudio.cloud`.



## Chapter 3

# StatPrep Annie

StatPrep Annie is a persona created to depict a real-world StatPREP instructor who is setting up their statistics course.

She's got a website for her course, a couple of interactive lessons, and so on.



Figure 3.1: StatPrep Annie



## Chapter 4

# Your course web site

As statistics instructors start using data in their classes, they find that they need to make data files available to students. An excellent way to do that is to put the files on a web site, so that the students can access them with a URL.

If your institution uses course support software such as Blackboard or Moodle, you may want to take advantage of those resources.

Many instructors don't have a web server available to them and aren't sure how to set up a web site. (And, warranted or not, many instructors grumble about Blackboard and Moodle) The point of this repository is to help you set up your own course web site on which you can place data files, etc. so that your students can easily get to them.

You don't need to know even what a "repository" is. You'll be able to add files to your web site and edit documents from an ordinary browser.


The main resource we'll use is *GitHub*. This is a free service that's very widely used by software engineers. We won't have to do any engineering, but you'll have to follow a few instructions.

At this point, make sure that you have a GitHub account. (See [@ref\(signing-up-for-cloud-services.html\)](#).)

### 4.1 Setting up your web site

1. Open up another browser window, so that you can look at these instructions at the same time as you are working on your site.
2. Login to GitHub in the browser window you opened in (1).
3. Open up one of the following links in the other browser window.
  - One-file web site uses only GitHub.
  - Many-file web site requires RStudio for editing. These are both "repositories" under the StatPREP organization.
4. Whichever option you selected in (2), find the "Fork" button in the upper right-hand corner of the StatPREP repository. Chances are, you'll be asked whether you want to set this up in your own account. You do.

At this point, you should have two browser windows open. One for these instructions, the other for your GitHub repository. You are going to be working in the GitHub repository window.

4. Press the  button. In the page that appears ...
  - a. Change the name of the repository to something appropriate for your course. Stat\_105? Keep it short and don't use any spaces. Use underscores instead.

## GitHub Pages

GitHub Pages is designed to host your personal, organization, or

✓ Your site is published at <https://dtkaplan.github.io/stat101/>

### Source

Your GitHub Pages site is currently being built from the `/docs` folder

master branch /docs folder ▾

Save

- b. Scroll down to the settings section headed GitHub pages. Select the drop-down menu to read “master branch /docs folder”.
  - c. Copy the link that follows “Your site is published at ...” This is the link you will give your students. Note that it is formed from your GitHub ID and the repository name. That’s why you want both of these to be memorable.
5. Try it out by pasting the link into the URL bar for another browser tab.
  6. Oh ... you might want to customize the content before you give the link to your students.

## 4.2 Customizing the site content.

How you are going to customize the content depends on whether you chose to create a one-page web site or a site that you can edit with RStudio.

### 4.2.1 One-page web site customization

When you are setting up your repository, you will be logged into GitHub and at a URL like this: `github.com/your_user_ID/your_course_name`.

When your students look at the repository, or when you make links to data files, etc., the URL will look like `your_user_ID.github.io/your_course_name`. Make sure it’s clear to you how the GitHub user URL differs from the URL for students.

We’re going to do some setup for your site, e.g. customizing the front page, adding data files, etc. So check that you are looking at your GitHub repository: `github.com/your_user_ID/your_course_name`.

You can see a list of files, starting with a folder called “docs”. The docs folder is where you will put all of the materials for your web site. Click on the name “docs” and you will see what files are already in the directory. There are two:

- `test.csv` - a really small CSV data file
- `index.md` - a text file containing the front page of your new site.

What might confuse you is that the site URL from the students’ point of view is something like `http://your_user_ID.github.io/your_course_name`, which does not include the word `docs`. Get used to it. The URL really does point to the `docs` directory. And, since there is a file called `index.md` in the `docs` directory, per the standard behaviour of web sites the contents of `index.md` are what will be displayed when someone points their browser to your `github.io` site.

Mostly, you’re going to do two things with your site:

1. Upload data files from your own computer into the `docs` folder on your site. Conveniently, there is an “Upload Files” button just for this purpose.

2. Edit the `index.md` file. To do this, click on the name `index.md`, which will open the file. You will see a little pencil icon; press that to edit the file. When you're done with your edits, scroll down and press the green "Commit changes" button. That simply saves your work. As soon as you've done this, the modified page is live on your web site, but it might take a few minutes and a refresh of your browser to see it.

#### 4.2.1.1 Putting links to data files on your own course web site

If you are going to use your site to provide student access to data sets of particular interest to you, you will want to put links and instructions on your course web site.

The markup that you include in your `index.md` file (in the `docs/` directory) might look like this:

```
## Google files used in class

- `Survey1 <- gs_read(gs_key("1ucevNh7wKLtOukyEpacUKi5_-KZUQGtIOONhWRnnnQ4"))`

## Data files

Data files for this week:

- `https://dtkaplan.github.io/stat101/test.csv`
```

To create the data table in your R session, copy and paste this command into your console:

```
```r
My_data <- read.csv("https://dtkaplan.github.io/stat101/test.csv")
```
```

#### 4.2.2 Customizing your site with RStudio

Outline:

- clone the repo
- open a new project in RStudio, choosing the option for a GitHub repository.
- Edit as needed. Every file you edit should be "knitted" to HTML.
- State, commit, push, and pull.



## Chapter 5

# Class data using Google Sheets

Collecting data interactively with students has several benefits:

1. Students see the connection between the data and their own actions. Students are especially motivated when the data is about them or something they are doing.
2. The inevitable imperfections in user-entered data serve as a lesson in coding factors and the measuring variables in a standard way.
3. The instructor (or individual students) can analyze the data and see how the analysis changes in real time, as more data rows are added.

### 5.1 Resources

Two example lessons, from the StatPREP 2018 Workshops are:

- Globe toss
- Riverboat card trick

You're welcome to use those interactive documents. Each document has a link to the spreadsheet for entering data. It's your own choice whether to start by clearing out any existing rows or add on to the rows from another session. **NOTE:** No guarantee that your class's data will be there later on, since someone else might have erased it. And so you might want to set up your own spreadsheet exclusively for your class's use.

The section below on *Setting up a new spreadsheet* shows how to create a spreadsheet for your class. Note that you should post the link to your spreadsheet on your course website, so that students can access it to enter data.

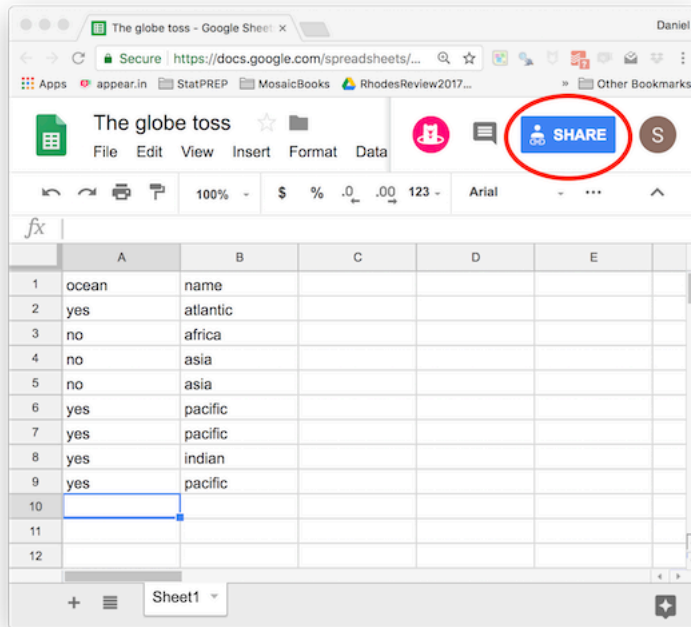
For data analysis, you have a choice of options:

1. Use the Read class data activity. The user will have to paste the command to read in your spreadsheet at the start of every command block. (See *Setting up a new spreadsheet*.)
2. Use your own R session. Again, you will need to give the command to read in your spreadsheet into the console.
3. Create your own tutorial document in Rmd. This presumes that you are comfortable editing Rmd documents and, if you want to give your students access, publishing them on a server. A working template is available via the StatPREP Workshops2018 project on [rstudio.cloud](https://rstudio.cloud). (Follow this link. to the file `Google_data_template.Rmd`.)

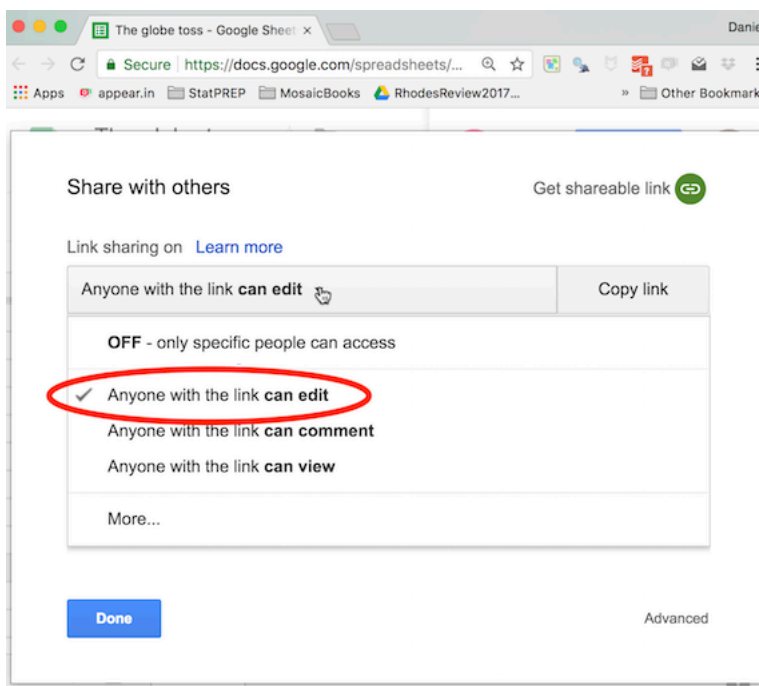
## 5.2 Setting up a new spreadsheet

You can modify this document to work with a spreadsheet of your own. Here's how.

1. Set up a Google spreadsheet. It's a good idea to populate it with some variable names and a few values. This will let you test to make things are working before you start the activity in class.
2. *Within* the Google spreadsheet document press the “Share” button.

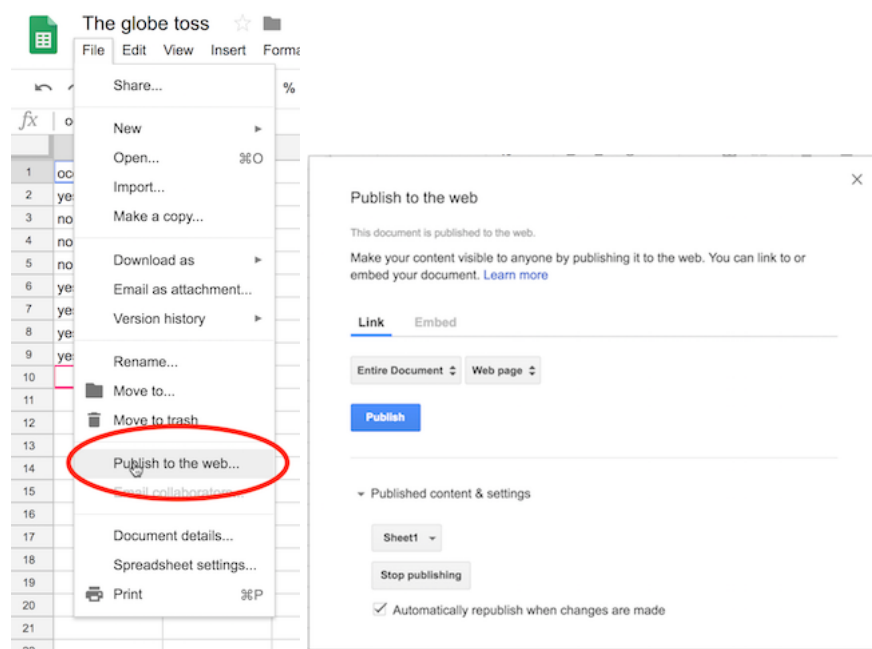


3. After you press “Share,” you will see a dialog box.





- Pull down the menu to select “Anyone with the link **can edit.**” (This is what lets your students add data.)
  - Copy the link and paste it somewhere you can get to it again. You’ll need it. The link will look like this: [https://docs.google.com/spreadsheets/d/1ucevNh7wKLtOukyEpacUKi5\\_-KZUQGtIOONhWRnnnQ4/edit?u](https://docs.google.com/spreadsheets/d/1ucevNh7wKLtOukyEpacUKi5_-KZUQGtIOONhWRnnnQ4/edit?u)
  - You will also need to copy the *key* that’s contained in the link. The key is just the central gibberish in the link, like this: **1ucevNh7wKLtOukyEpacUKi5\_-KZUQGtIOONhWRnnnQ4**
4. Put the link on your course web site so that your students can get to it. That’s how they will access the spreadsheet for entering data.
  5. Back in your Google sheet, select the File/Publish\_to\_the\_web menu item. Use the resulting dialog box to publish the entire document.



6. Create the R command that will load the spreadsheet data into your R session.

```
Globe <- gs_read(gs_key("1ucevNh7wKLtOukyEpacUKi5_-KZUQGtIOONhWRnnnQ4"))
```

In forming the command, replace the quoted string in the above with your own key. The key is located in the center of the link to the spreadsheet: an incomprehensible set of characters similar to that highlighted in **bold** in the example in step (3). You might also replace **Globe** with a name that’s more suited to your own activity.

Paste the command you’ve created someplace handy. You’ll need it. (Suggestion: Paste it next to the link in step (4).)

Time for class!

7. Once you reach the point in your class where you want to do statistics on your data, bring up the lesson document provided for this purpose by StatPREP located here. That document has several R command chunks, all of which are blank. You can put any R commands in those chunks, **but make sure** that the command from step (5) always is the first command in any chunk that you use. That way, whenever you run the code in the chunk, the data will be read in from Google. Keep in mind that the chunks are all independent of one another, so you’ll need to read in the data in any chunk you use.

Try it out in the following command chunk:

- a. Paste in the command you created in step (5).

- b. Below that, add any R commands you like.

For instructors who want to write their own tutorial, you will find that this simplifies things since you can arrange to have the spreadsheet data read-in globally and not have to put the data-reading command in every chunk. Use the template `.Rmd` document provided by StatPREP [here](#). Modify the chunk named `read_data` the top of the document by inserting your own command (with your own key). Notice that in any new chunk you create, you'll have to reference the `read_data` chunk as the `exercise-setup`. The chunks already in the template document do this, so you can just copy (and rename!) an existing chunk.

## Chapter 6

# Using RStudio.cloud

Just some preliminary notes ...

Under preferences/Rmd, arrange to have the preview opened in the Viewer tab. It doesn't seem to work to leave it as opening in a web page.