

# RSCH 630 - Research Methods

Final Exam

*Your Name*

*Due 17:00 6/7/2016*

## Abstract

“The purpose of this entire course has been to introduce techniques that can help you during your research. And while it’s easy to focus on the mathematical methods, you’re likely to spend more time on the grunt-work tasks, such as getting and cleaning data, building tables and figures, numbering equations and citing sources. On the first day of class I stated that a lot of effort has gone into developing tools that make these tasks simpler - and in some cases almost trivial. It would be wrong for me to simply make such a statement without showing you how to take advantage of these tools. Therefore, the purpose of this exam is to have each of you build (and use) a data science toolkit that links these tools together. R/RStudio will comprise the core of this toolkit. However, even if you don’t use R, this toolkit will allow you to add in almost any other language you want. Building the toolkit will take some effort up-front and will introduce you to several new concepts, but once it’s up and running the workflow becomes almost seamless. Alright, let’s get started.”

## Installing Resources

### Step 1 - Create a GitHub Account and Fork the rsch-630 Repo

- Go to GitHub and sign up for an account
  - Choose a username
  - Add an email address (don’t use your afit.edu address as your primary, you can add it as a secondary address later)
  - Enter a password
- GitHub is a version control tool for ASCII or UTF-8 encoded files
  - Organizes code, text files, images, etc. into repositories ‘repos’
  - Easy to collaborate on projects with other users any where in the world without

- No more emailing files, DVD's, USB drives - It's like dropbox on steroids
- It can be used on AF NIPR machines (probably not all, but haven't found one yet that couldn't)
- GitHub has a learning curve, but once you get the hang of it you don't go back
- Outside of AFIT/AF, GitHub is the standard - some employers prefer GitHub profiles over resumes

“It's become so important to its 13 million users that if GitHub goes down, the software development world practically stops.” [1]

## Step 2 - Install R

On personal machines you can install R version 3.3.0 (2016-05-03) from <https://cloud.r-project.org/> selecting the default installation

On AFIT machines you can install R version 3.3.0 (2016-05-03) from the R:/Stats/R folder

- Select the green installer icon to ensure R installs to Program Files
- Follow the prompts for a default installation

## Step 3 - Look Around

Open R up... It's ugly = no one would use R if this was all there was to it. Thankfully there's a **LOT** more

## Step 4 - Install a Bunch of R Packages

- `install.packages(c('installr', 'devtools', 'rmarkdown'))`
- `install.packages(c('shiny', 'RJSONIO', 'xtable'))`
- `install.packages(c('DT', 'knitcitations', 'RefManageR'))`

## Step 5 - Install RStudio

- If you using a Windows machine this is really easy, just run `installr::install.rstudio()`
- If you're not using Windows you'll have to go to <https://www.rstudio.com/products/rstudio/download/> to install the most recent version
- **NOTE:** RStudio is an IDE
  - You can work on files written in different languages side-by-side
  - You can mix languages together in a single file

## Step 6 Install & Connect Rtools

If you using a Windows machine this is again really easy, just run `installr::install.rtools()`

If you're not using Windows you'll have to go to <https://cran.r-project.org/bin/windows/Rtools/> to install the Rtools 3.3

## Step 7 Install & Connect Git

If you using a Windows machine, guess what? Just run `installr::install.git()`

If you're not 2 using Windows go to <https://git-scm.com/>

## Working Example

### Scraping data from Data.gov

```
[1] "httr" [1] "RSocrata" [1] "rdatacite" [1] "data.table"
```

## Processing The Data

Inserting tables automatically

## References

[1] J. Bort, “Why \$2 billion startup GitHub is apparently in crisis, again,” *Business Insider*. Feb-2016.

[2] T. Fliedner and J. Liesiö, “Adjustable robustness for multi-attribute project portfolio selection,” *European Journal of Operational Research*, vol. 252, no. 3, pp. 931–946, Aug. 2016.

Table 1: Extracranial Procedures Medical Payment Information

Provider State	Covered Charges	Medicare Payments	Total Discharges
AK	34805.13	14815.73	23.00
AL	737022.05	237893.71	879.00
AR	439412.21	176920.66	652.00
AZ	802640.65	316425.33	606.00
CA	3758360.36	1022881.47	1715.00
CO	352522.07	124718.52	255.00
CT	324480.09	220795.03	384.00
DC	142109.63	52814.37	47.00
DE	71899.56	38542.59	144.00
FL	3453570.67	896814.74	2847.00
GA	817751.40	335019.92	1015.00
HI	27809.95	17828.91	24.00
IA	308569.52	157788.85	518.00
ID	115049.71	72258.57	113.00
IL	1775623.21	663680.82	1240.00
IN	967990.18	421914.42	1078.00
KS	413753.56	175982.37	594.00
KY	402152.95	209508.20	728.00
LA	854004.03	286390.00	743.00
MA	571138.89	390889.57	847.00
MD	202294.89	359042.13	706.00
ME	43441.97	37621.12	205.00
MI	892884.86	519228.72	1325.00
MN	351564.46	188878.79	445.00
MO	783848.97	342104.18	1030.00
MS	412137.08	170357.37	588.00
MT	90997.88	58351.72	174.00
NC	571979.55	322140.07	1141.00
ND	86702.08	70695.29	178.00
NE	214240.29	111884.31	318.00
NH	193600.00	100773.88	216.00
NJ	1088217.96	415506.82	819.00
NM	227310.18	79553.73	128.00
NV	522703.23	137947.50	311.00
NY	1069991.17	694620.36	1336.00
OH	1352733.33	564024.31	1186.00
OK	239155.88	117278.91	463.00
OR	295619.25	154057.80	354.00
PA	1459743.61	536073.30	1003.00
RI	74741.41	43908.66	76.00
SC	643419.21	228318.54	721.00
SD	87347.69	49616.82	156.00
TN	845356.01	352225.32	986.00
TX	2378767.25	900469.38	2759.00
UT	123586.15	69379.32	109.00
VA	120592.55	65796.73	161.00
VT	17934.35	17226.83	57.00