

Quantified Grad Student

Derek Nedveck // nedveck@gmail.com

6/26/2014

What I'm going to present

- ▶ Intro - why did I do this
- ▶ Data collection
- ▶ Analysis
- ▶ Future directions - other questions?

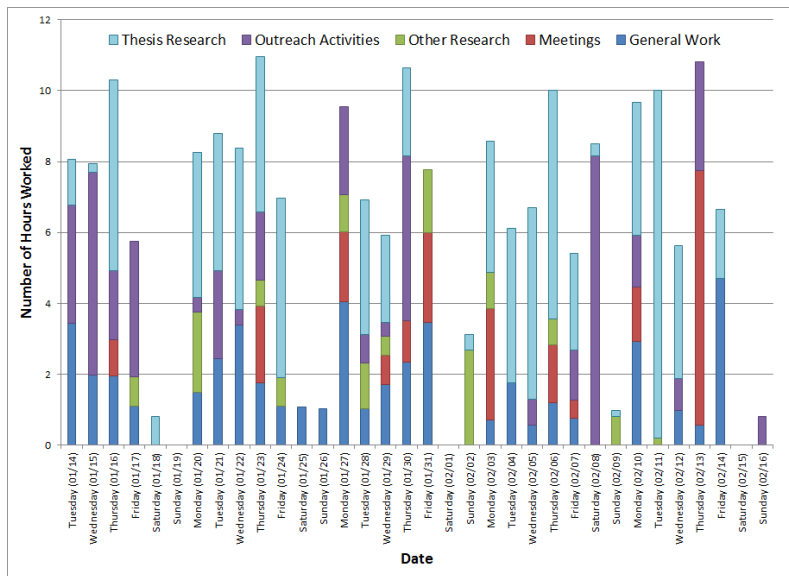
Motivation

- ▶ Entering grad school - how much time do I need to invest?
- ▶ Anecdotal: it will take all of your time
- ▶ Lots of claims, but no data

Questions

- ▶ How much time do I invest in Grad School?
 - ▶ How much time do I spend on campus?
 - ▶ How much time am I working on grad school stuff?
 - ▶ What percent of time that I am on campus am I actually getting work done?

One data point



<http://rddenton.blogspot.com/2014/02/who-works-80-hours-week-in-academia.html>

Who am I?

Grad student in Plant Biology

- ▶ Genomics (computational work), and local adaptation (field / greenhouse / lab work)
- ▶ Advisor says I'm making good progress
- ▶ How do I compare to other Grad students?
 - ▶ I'm probably not very representative... but then what is a "normal" grad student

Data Collection

Tools

- ▶ Android App: Gleeo
 - ▶ Time doing work
- ▶ Google Spreadsheet
 - ▶ Time I'm on campus

How do I categorize my time?

How many categories?

- ▶ Tracked everything in a spreadsheet for a week
- ▶ Partition the tasks into categories, with the least amount going into a “misc” bin
- ▶ Still iterating, implement another scheme starting fall semester

Analysis

Packages

- ▶ lubridate to work with dates, times, and intervals
- ▶ dplyr to work with dataframes
- ▶ ggplot2 to make all the graphics

lubridate is pretty handy

```
# getting the week number from the date
clock.df$week <- week(clock.df$date)
# Using intervals, dividing the seconds in the
# interval by the exact seconds in an hour
clock.df$campustime <- interval(
  ymd_hms(paste(clock.df$date,
                 clock.df$in_time)),
  ymd_hms(paste(clock.df$date,
                 clock.df$out_time))
) / ehours(1)
```

dplyr is neat

```
clock.df %>% group_by(day) %>%  
              summarize(mean(campustime))
```

```
## Source: local data frame [5 x 2]
```

```
##
```

```
##      day mean(campustime)
```

```
## 1   Mon           9.277
```

```
## 2  Tues           9.449
```

```
## 3   Wed           8.721
```

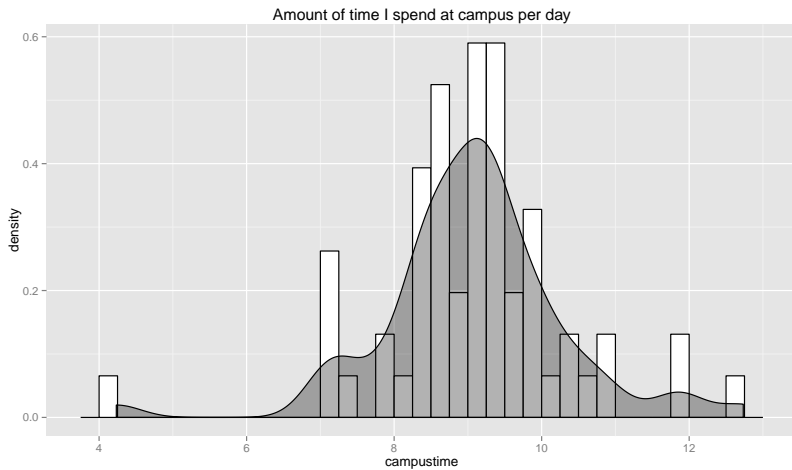
```
## 4 Thurs           9.645
```

```
## 5   Fri           8.315
```

Campus Time

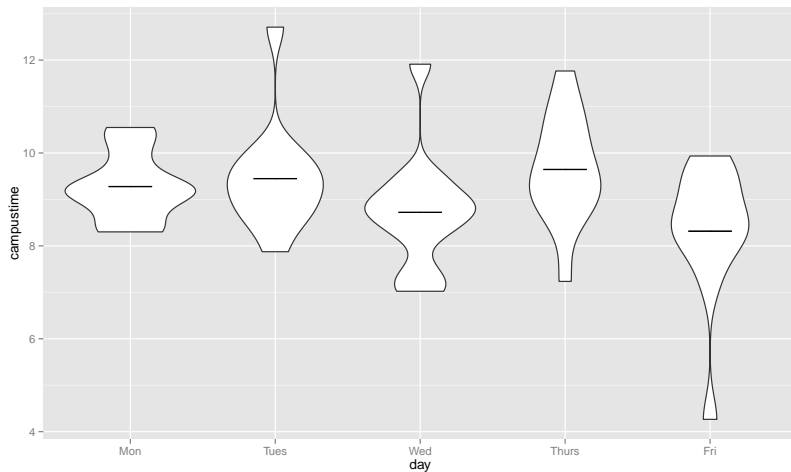
- ▶ Time that I am at my office / on campus each day, Monday - Friday

How long am I at campus?

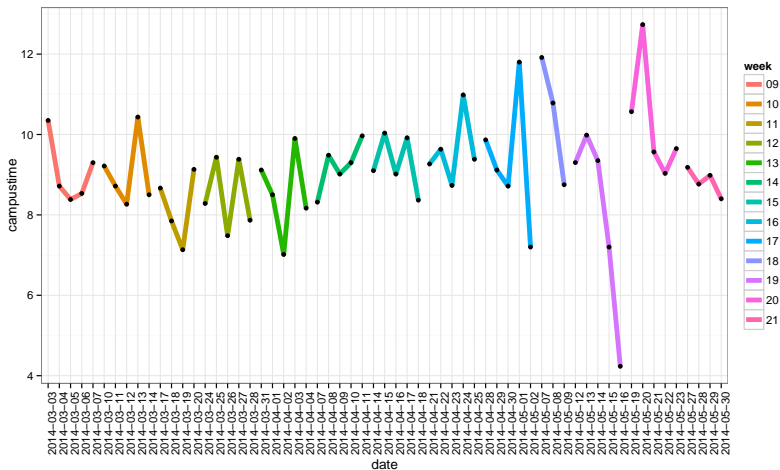


```
##   Min.   Mean   Max.
##  4.23   9.08 12.70
```

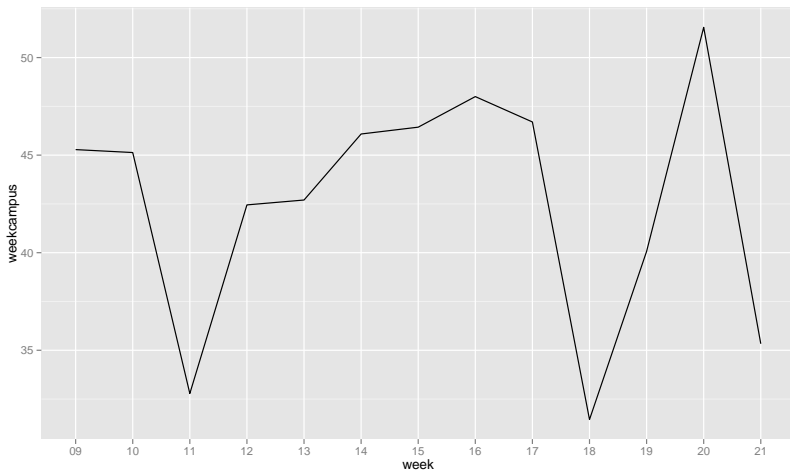
Grouped by weekday



Time at campus varies over time



Time per week



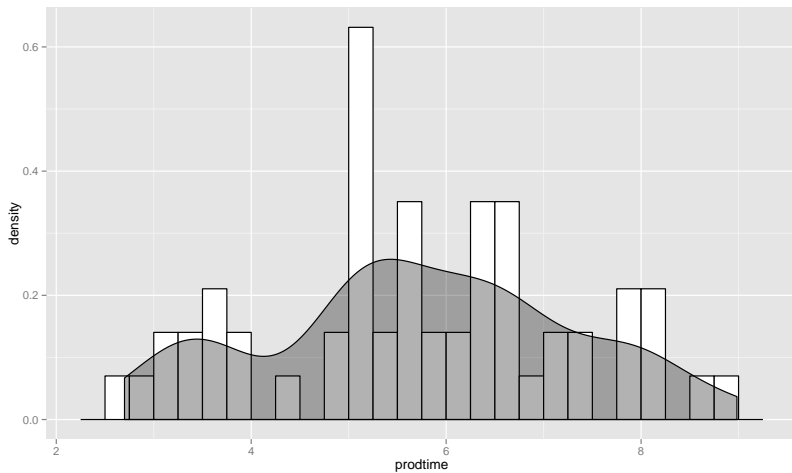
Min. Mean Max.

31.4 42.6 51.6

Productive Time per day

$$ProductiveTime = TotalTrackedTime - PersonalTrackedTime$$

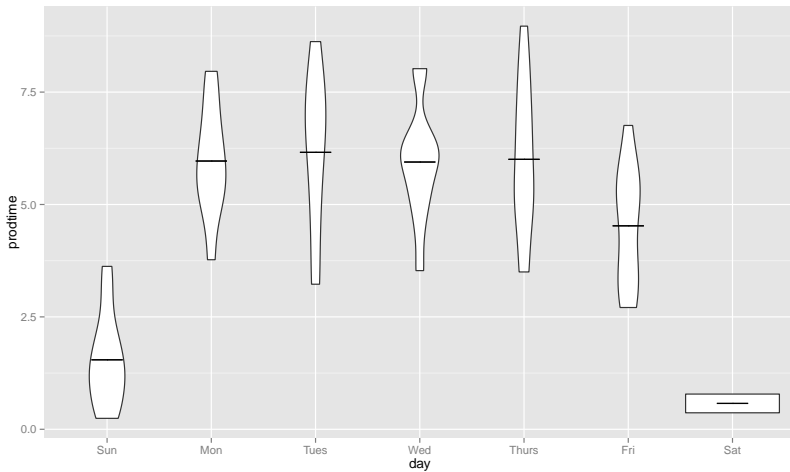
Distribution of Productive Time per day



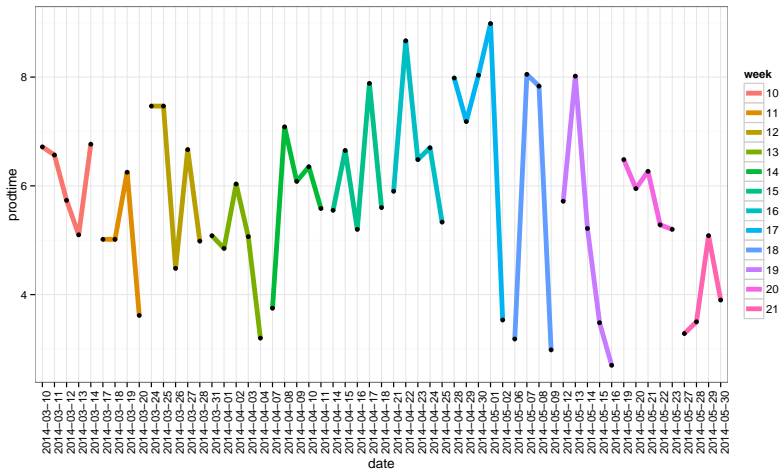
Min. Mean Max.

2.70 5.73 8.98

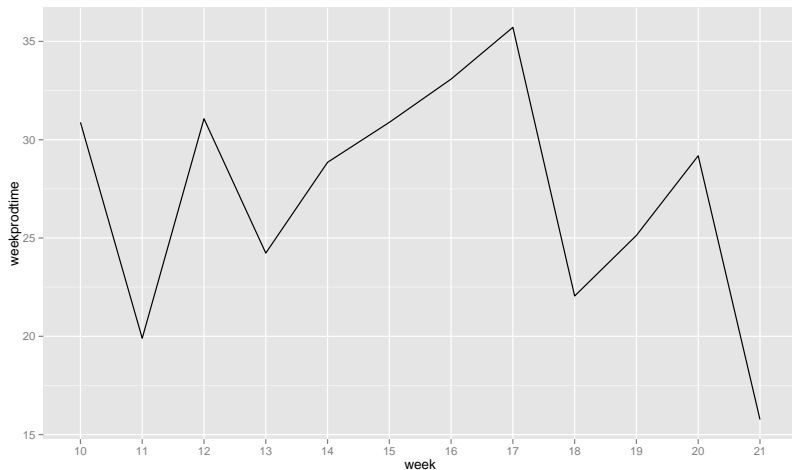
Distribution of Productive Time Each Day



Productive Time varies over time



Weekly Productive Time



Min. Mean Max.

15.8 27.2 35.7

Efficiency

- ▶ When I am at campus, how much time of that am I actually working?

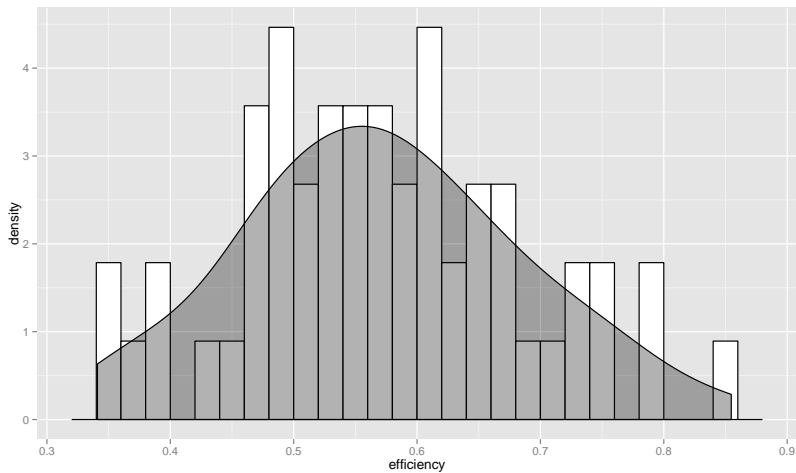
$$Efficiency = \frac{TotalTrackedTime - PersonalTrackedTime}{TimeAtCampus}$$

lubridate is really handy here

My solution isn't elegant. . .

- ▶ make the interval that I am at campus, one per day
- ▶ for each tracked task, make an instant
- ▶ for loop to go by each campus interval and find which instants are within that interval, store in df
- ▶ populate a new df with only those instants in the interval
 - ▶ for loop, and rbind

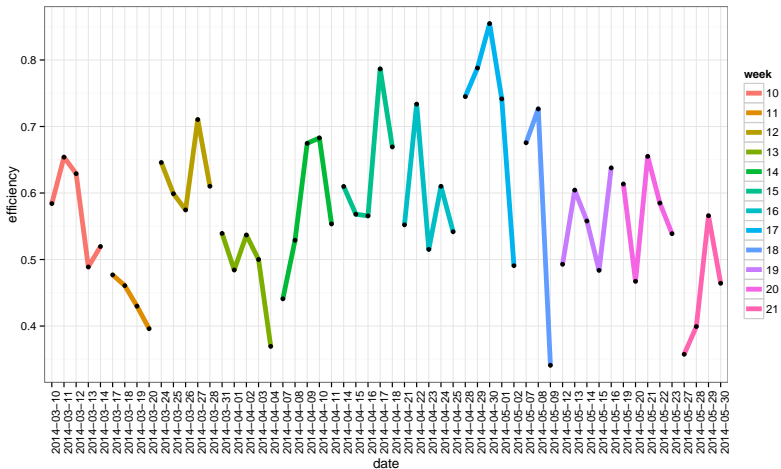
Distribution of my efficiency



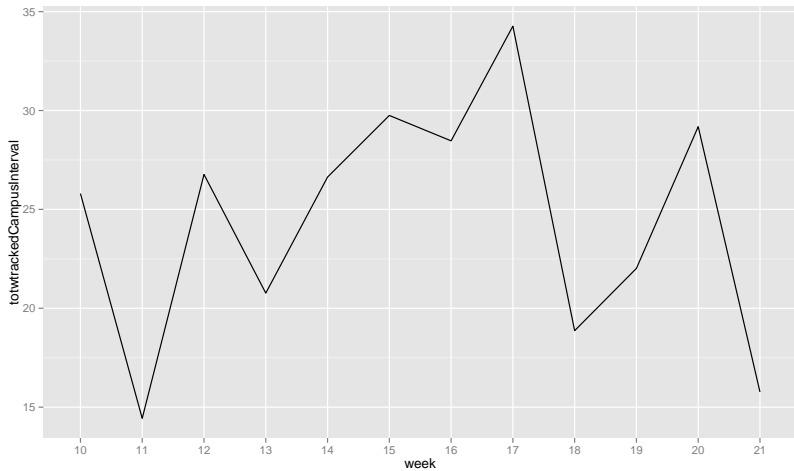
Min. Mean Max.

0.341 0.572 0.855

Efficiency vary over time



Efficiency vary by week



In Summary

How much time do I invest in Grad School?

- ▶ I spend a mean of **9.08 hrs per day**, and **42.6 hrs per week** at campus
- ▶ I work a mean of **5.73 hours per day**, and **27.2 hours per week**

How productive am I?

- ▶ My mean productivity is **0.572** hrs worked / hr at campus

Project Data

- ▶ I haven't used my categorized tracked time data yet
- ▶ next goal: make a graph of time invested in projects as a function of date

Self Improvement?

- ▶ I've answered the questions I started with, but how can I use this to improve my studies?
- ▶ Act of tracking improving my focus?
- ▶ Guilt of bad data?

Advisor says:

"... issues with Brownian Motion in accomplishing your projects. . . "

Future work

- ▶ Cal Newport and Deep Thought — is there a way to track this?
- ▶ How much do grad students read?
 - ▶ I can kind of answer this already

More Data and Passive tracking

- ▶ Moves app to potentially track my campus time, also time at gym
- ▶ Using Rescue Time to track computer usage
 - ▶ especially quick email usage — often not tracked
- ▶ Pull in weather data?
- ▶ What things can affect my productivity, and is it easy to collect these data?

This is on GitHub

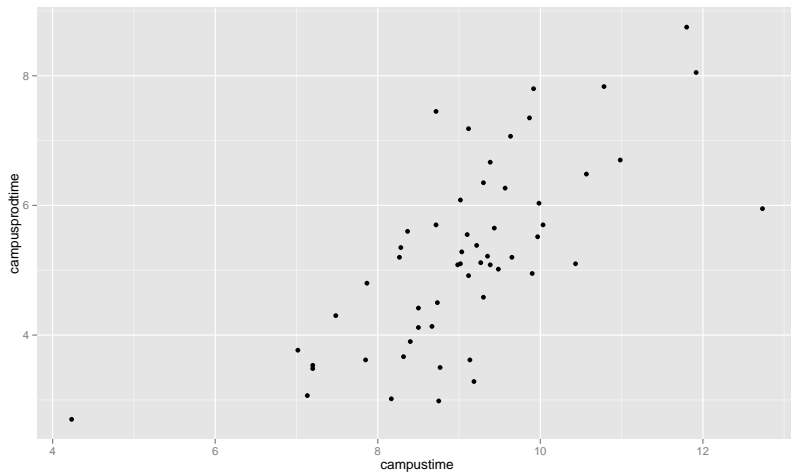
<https://github.com/dnedveck/qgs>

- ▶ `./TCRUG/TCRUGpres.Rmd`
 - ▶ This presentation
- ▶ `./*_GleeeoExport.csv`
 - ▶ Time tracking data
- ▶ `./*_clocktime.csv`
 - ▶ Campus time data

Any Questions?

- ▶ nedveck@gmail.com
- ▶ dnedveck.com
- ▶ <https://github.com/dnedveck/qgs>

Productive time at campus given time at campus



Efficiency given campus time

