R Notebook: Workspace Audit for [jessica.legarza]

Contents


```
library(devtools);
```

```
## Warning: package 'devtools' was built under R version 4.0.3
```

```
## Loading required package: usethis
```

```
library(humanVerseWSU);

path.github = "https://raw.githubusercontent.com/MonteShaffer/humanVerseWSU/master/";

include.me = paste0(path.github, "misc/functions-nlp.R");
source_url( include.me );
```

```
## SHA-1 hash of file is 704afa69d52215d315cb5f59cdc020b0bbfd0b13
```

```
## Warning: package 'tm' was built under R version 4.0.3
```

```
## Loading required package: NLP
```

```
## Warning: package 'NLP' was built under R version 4.0.3
```

```
## Warning: package 'SentimentAnalysis' was built under R version 4.0.3
```

```
##
## Attaching package: 'SentimentAnalysis'
```

```
## The following object is masked from 'package:base':
##
##     write
```

```
include.me = paste0(path.github, "misc/functions-nlp-str.R");
source_url( include.me );
```

```
## SHA-1 hash of file is 6bdb234fa84eea995969dc29d6ff2a78f3982131
```

```
include.me = paste0(path.github, "misc/functions-nlp-stack.R");
source_url( include.me );
```

```
## SHA-1 hash of file is 034efbce0405954198545f8798e119b77a4809c9
```

```
include.me = paste0(path.github, "misc/functions-nlp-pos.R");
source_url( include.me );
```

```
## SHA-1 hash of file is d8c8cf01c8ead1b6d4228891aa52bac77084a6e7
```

```
## Warning: package 'openNLP' was built under R version 4.0.3
```

```
include.me = paste0(path.github, "humanVerseWSU/R/functions-encryption.R");
source_url( include.me );
```

```
## SHA-1 hash of file is da71dde620bed33db055778b752eefb476f7bf6b
```

```
###### UPDATES TO dataframe subset function ######
# inflation adjustments for NA ... and improvements on subsetting
include.me = paste0(path.github, "humanVerseWSU/R/functions-str.R");
source_url( include.me );
```

```
## SHA-1 hash of file is 5f57d466d946285783f15bb3b4d97ca48a951d0b
```

```
include.me = paste0(path.github, "humanVerseWSU/R/functions-file.R");
source_url( include.me );
```

```
## SHA-1 hash of file is 091aefe67c59f7dc1644c268fdd0b0a183c525ef
```

```
include.me = paste0(path.github, "humanVerseWSU/R/functions-dataframe.R");
source_url( include.me );
```

```
## SHA-1 hash of file is 1149cbf3e865f692b50d4d1983e6364dc56ce62d
```

```
include.me = paste0(path.github, "humanVerseWSU/R/functions-inflation.R");
source_url( include.me );
```

```
## SHA-1 hash of file is b6d29327e3fe030ca132b135f4a89b6fc6a61a66
```

Workspace Audit

You should place this file in your github local path for this course, in a subfolder called "final" ... for me that is `C:/_git_/WSU_STATS419_FALL2020/final/`;

Set your WSU id and location of your github USERNAME/REPO

```
wsu.id = "jessica.legarza"; # change this to yours.
github.repository = "jsmith0434/WSU_STATS419_FALL2020";
# be sure to change the notebook title at the top of this document to be "you" based on your wsu.id
```

Code Frequency on GitHub

Browse to https://github.com/MonteShaffer/WSU_STATS419_FALL2020/graphs/code-frequency where you replace my repository name with yours. Screenshot and include the graphic in this folder, named "github-code-frequency.png".

Source: GitHub

Commit Frequency on GitHub

Browse to https://github.com/MonteShaffer/WSU_STATS419_FALL2020/graphs/commit-activity where you replace my repository name with yours. Screenshot and include the graphic in this folder, named "github-commits.png".

Source: GitHub

General Workspace Environment Data

```
path.to.git.local = "C:/_git_/WSU_STATS419_FALL2020/";
setwd(path.to.git.local);

path.to.git.final = paste0(path.to.git.local,"final/");
```

General Workspace Environment Data

```
my.object = list();

sg = Sys.getenv();
  snames = names(sg);
```

```
  svals = as.character(sg);
sgdf = as.data.frame(cbind(snames,svals));

my.object$environment = sgdf;

sw = Sys.which(c("ftp", "ping", "texi2dvi", "this-does-not-exist", "make", "latex", "pdflatex", "luatex
  snames = names(sw);
  svals = as.character(sw);
swdf = as.data.frame(cbind(snames,svals));

my.object$bin = swdf;
my.object$packages = as.data.frame(installed.packages());

my.object$local.dirs = list.dirs(path.to.git.local,recursive = FALSE);
my.object$local.files = list.files(path.to.git.local,recursive = TRUE);
```

Local Computer Setup

Dropbox

Using your file-explorer, browse to your "Dropbox" folder on your computer, enter the "**student_access**" folder for this course, and take a screenshot. Place in this folder "final/dropbox.png" ...

Source: LOCAL COMPUTER

Git tool

Take a screenshot of your gitbash tool (or if you use Github Desktop, that will do). Be certain the screenshot is referencing your repository for this course. Place in this folder "final/git.png" ...

Source: LOCAL COMPUTER

Git History

Open your git tool from the command line and browse to the correct folder ... Type `git status` to verify you are in the right place.

Now type: `git log --shortstat --pretty=format:'%h|%t|%p|%ai|%ae|%an|%ci|%ce|%cn|%N|%f|%s|%b' | paste - - - > final/git-history.log` which will create a file with your git history as a text file. We will import and perform a basic graph.

```
#  git log --shortstat --pretty=format:'"%h","%an","%ae","%aD","%s",' | paste - - - > final/git-history
# git log --oneline --decorate > final/git-history.log
# git log --shortstat --pretty=format:'%h|%t|%p|%ai|%ae|%an|%ci|%ce|%cn|%N|%f|%s|%b' > final/git-histor
#####################################################
############ BELOW IS THE FINAL ANSWER ############
#####################################################
# git log --shortstat --pretty=format:'%h|%t|%p|%ai|%ae|%an|%ci|%ce|%cn|%N|%f|%s|%b' | paste - - - > fi

# https://mirrors.edge.kernel.org/pub/software/scm/git/docs/git-log.html#_pretty_formats
```

```
my.git = list();

correct_path = "C:\\Users\\jsmit\\Desktop\\WSU\\DataAnalytics\\STAT419\\WSU_STATS419_FALL2020\\final\\"
raw.df = read.csv(paste0(correct_path,"git-history.log"), header=FALSE, sep="|", quote="");
colnames(raw.df) = c("hash","tree","parent","author.date","author.email","author.name","commit.date","c
#raw.df;

total.commits = nrow(raw.df);



date.strings = raw.df$commit.date;
date.types = c("%m","%d","%W","%j","%u","%H");
date.names = c("month", "day", "year.week", "year.day", "day.week","day.hour");

time.df = convertDateStringToFormat(date.strings, date.types, date.names);
#time.df;

changes.df = NULL;
for(i in 1:total.commits)
  {
  my.commit = removeWhiteSpace(strsplit(raw.df$content.b[i],",",fixed=TRUE)[[1]]);

  my.detail = strsplit(my.commit," ",fixed=TRUE);

  row = c(0,0,0);
  for(j in 1:length(my.detail))
    {
    val = as.numeric( my.detail[[j]][1] );
    row[j] = val;
    }
  changes.df = rbind(changes.df,row);
  }
```

Git Log Data

```
## Warning: NAs introduced by coercion

## Warning: NAs introduced by coercion

## Warning: NAs introduced by coercion

## Warning: NAs introduced by coercion

## Warning: NAs introduced by coercion

## Warning: NAs introduced by coercion
```

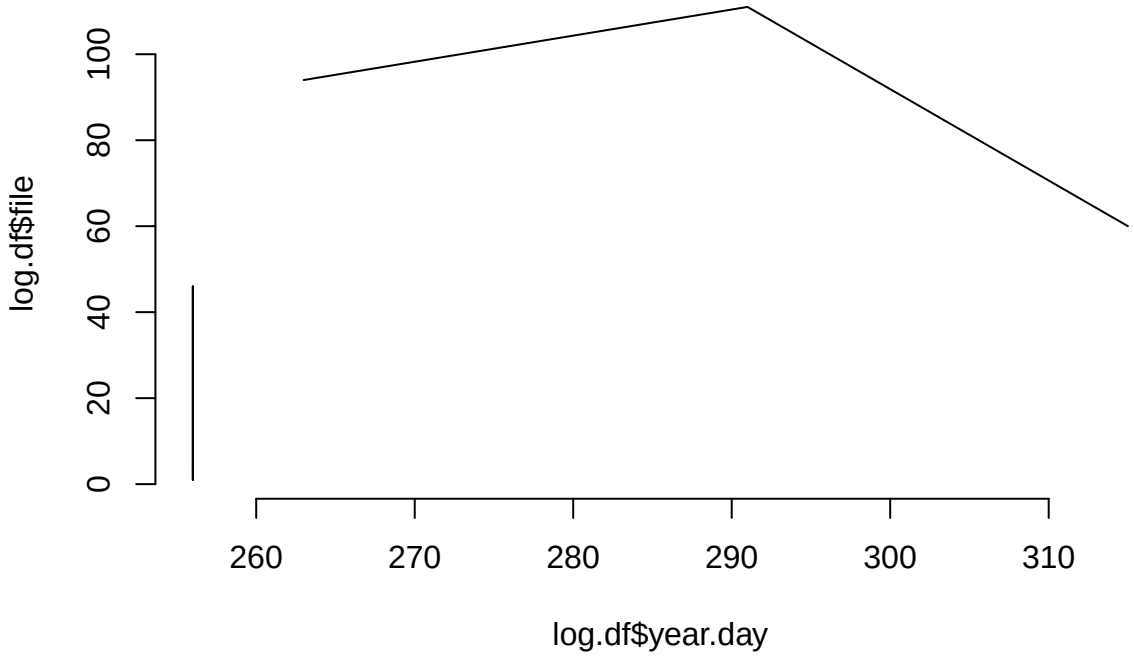
```
changes.df = as.data.frame(changes.df);
colnames(changes.df) = c("file","insert","delete");
#changes.df;
```

```
log.df = cbind(raw.df[,c(1:3,7:9,12)], time.df, changes.df);

my.git$raw.df = raw.df;
my.git$total.commits = total.commits;
my.git$time.df = time.df;
my.git$changes.df = changes.df;
my.git$log.df = log.df;

log.df;
```

```
##           hash    tree  parent               commit.date         commit.email
## row    1c11f0d d4b3d86 d0c2c85 2020-11-10 11:03:39 -0800 jsmith0434@yahoo.com
## row.1  d0c2c85 c72f5de 1fb87b5 2020-10-17 08:41:03 -0700 jsmith0434@yahoo.com
## row.2  1fb87b5 d02b9a8 13aec46 2020-09-19 07:10:01 -0700 jsmith0434@yahoo.com
## row.3  13aec46 84aefd4 9f42d16 2020-09-13 15:32:48 -0700   noreply@github.com
## row.4  9f42d16 35cd3d9 0352d18 2020-09-13 15:32:11 -0700   noreply@github.com
## row.5  0352d18 62210c2 d04ae9e 2020-09-13 15:31:00 -0700   noreply@github.com
## row.6  d04ae9e 655daa9 57d72ab 2020-09-13 15:30:29 -0700   noreply@github.com
## row.7  57d72ab c417623 8daba4c 2020-09-13 15:30:00 -0700   noreply@github.com
## row.8  8daba4c 244207a 82db06e 2020-09-13 15:29:41 -0700   noreply@github.com
## row.9  82db06e 9f5f62e 376f07f 2020-09-12 12:47:08 -0700 jsmith0434@yahoo.com
## row.10 376f07f bf2c310 54dce5e 2020-09-12 09:49:29 -0700 jsmith0434@yahoo.com
## row.11 54dce5e 4d53749 fd0042e 2020-09-12 09:47:54 -0700 jsmith0434@yahoo.com
## row.12 fd0042e 530628f 081c001 2020-09-12 08:11:48 -0700 jsmith0434@yahoo.com
## row.13 081c001 a22663b         2020-09-12 07:33:14 -0700   noreply@github.com
##          commit.name                                    content.s month day
## row    Jessica Smith                  pushing files for the project    11  10
## row.1  Jessica Smith   pushing new and updated instructor notebooks    10  17
## row.2  Jessica Smith                              week04 assignment     9  19
## row.3         GitHub                          Delete datasetsHW.pdf     9  13
## row.4         GitHub                          Update datasetsHW.Rmd     9  13
## row.5         GitHub  Delete Jessica Smith_Writing Assignment 1.pdf     9  13
## row.6         GitHub Delete Jessica Smith_Writing Assignment 1.docx     9  13
## row.7         GitHub   Delete Jessica Smith_Reading Assigment 1.pdf     9  13
## row.8         GitHub  Delete Jessica Smith_Reading Assigment 1.docx     9  13
## row.9  Jessica Smith                      updated measurements data     9  12
## row.10 Jessica Smith                                libraries-again     9  12
## row.11 Jessica Smith                                     libraries     9  12
## row.12 Jessica Smith                           personality dataset     9  12
## row.13        GitHub                                Initial commit     9  12
##        year.week year.day day.week day.hour file insert delete
## row           45      315        2       11   60 134771    137
## row.1         41      291        6        8  111  49181     17
## row.2         37      263        6        7   94  11904  38946
## row.3         36      257        7       15   NA      0      0
## row.4         36      257        7       15   NA      1      2
## row.5         36      257        7       15   NA      0      0
## row.6         36      257        7       15   NA      0      0
## row.7         36      257        7       15   NA      0      0
## row.8         36      257        7       15   NA      0      0
## row.9         36      256        6       12    3     20      9
## row.10        36      256        6        9    1      1      0
## row.11        36      256        6        9    3   1881      0
```

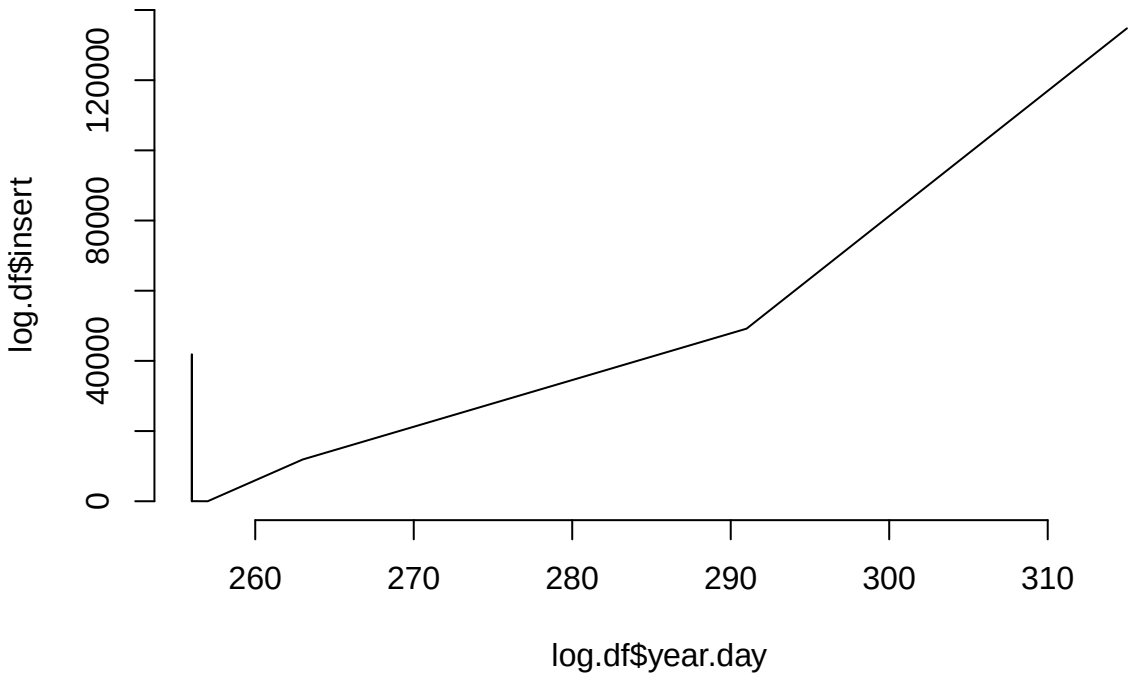
```
## row.12        36      256        6        8   46  41812      0
## row.13        36      256        6        7    3     61      0
```

```
plot(log.df$year.day, log.df$file, bty="n", type="l");
```

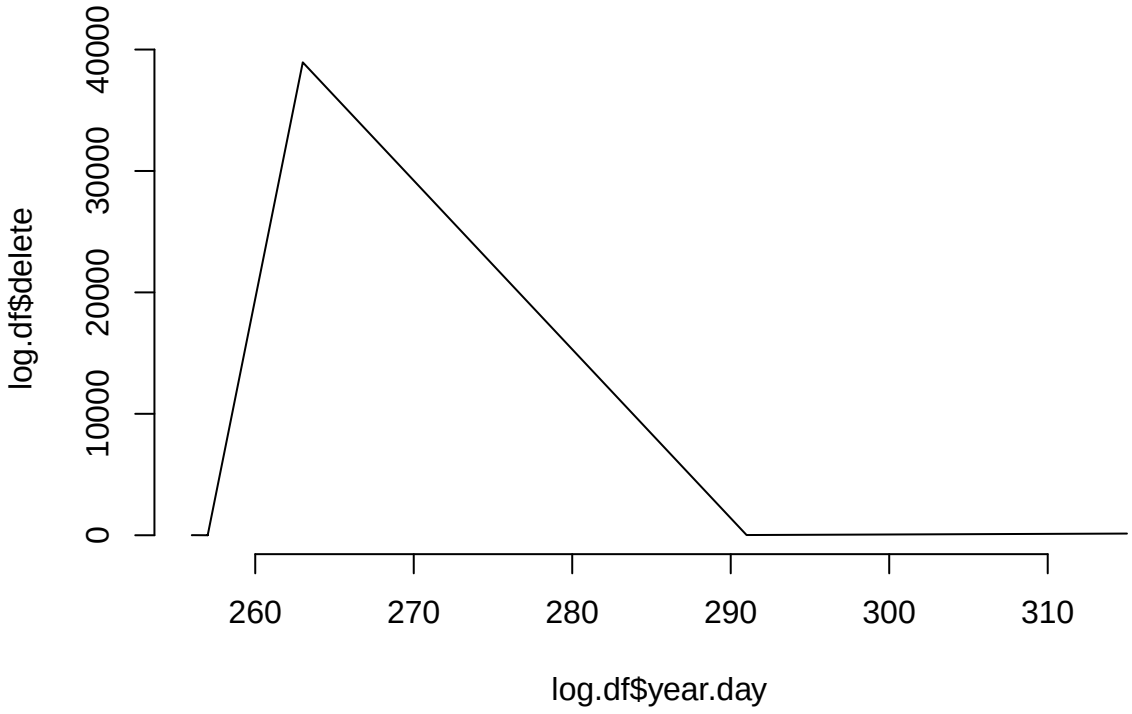


Git Log Graphics

```
plot(log.df$year.day, log.df$insert, bty="n", type="l");
```

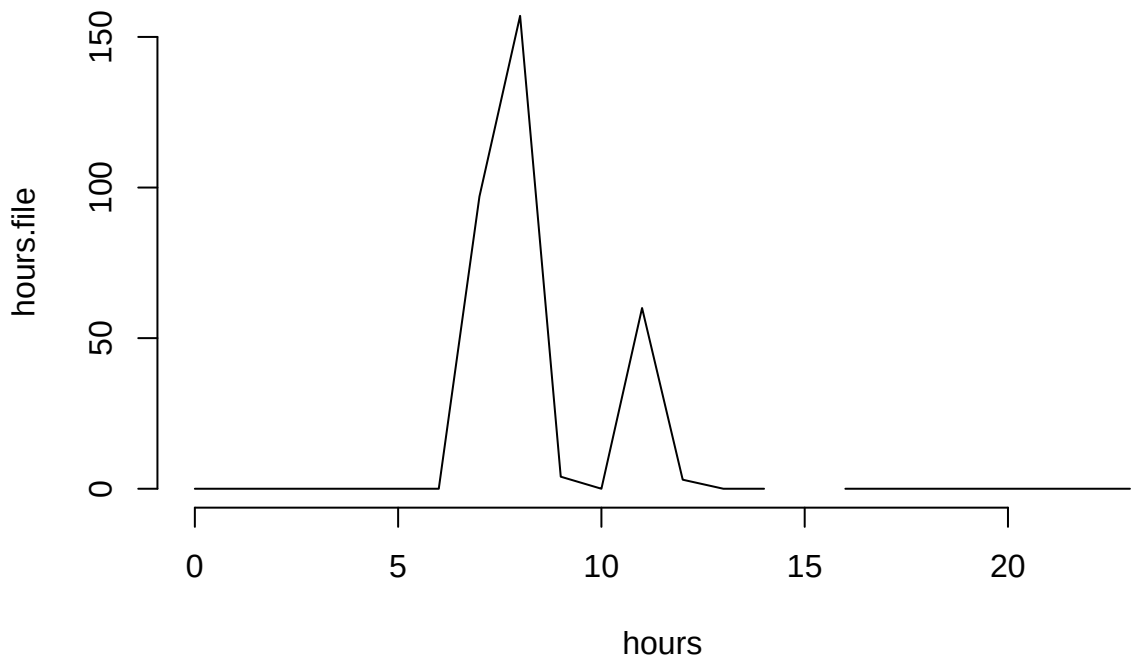



```
plot(log.df$year.day, log.df$delete, bty="n", type="l");
```

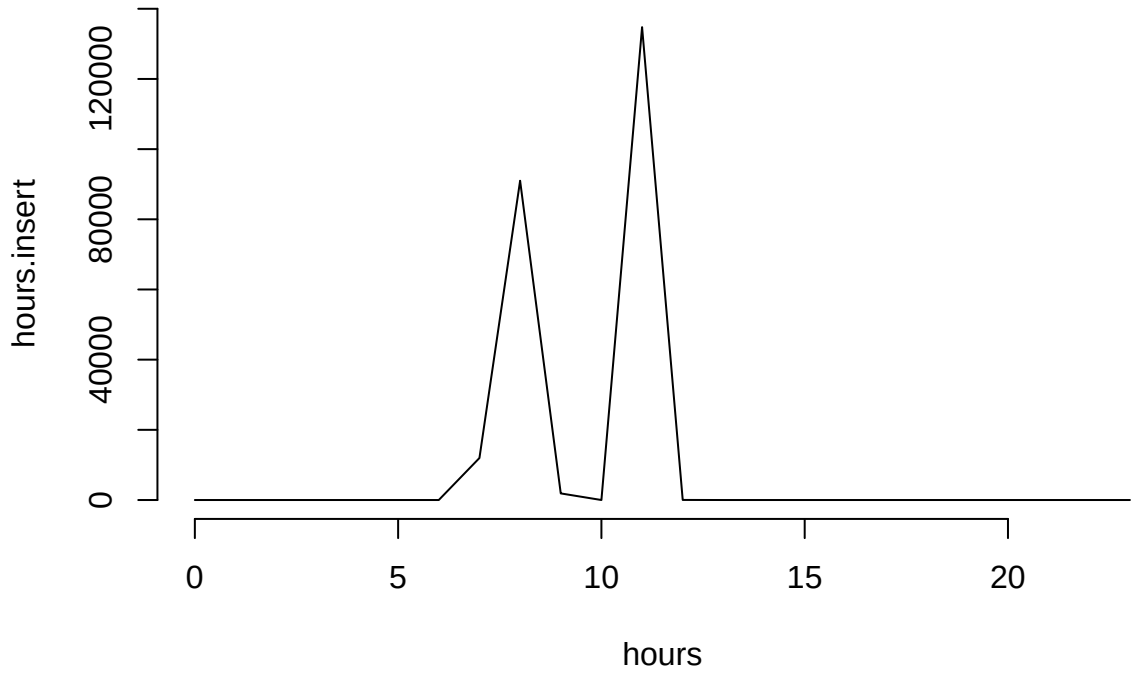



```
weeks = sort(unique(log.df$year.week));
n.weeks = length(weeks);
weeks.file = weeks.insert = weeks.delete = numeric(n.weeks);

days = 1:7;
n.days = length(days);
days.file = days.insert = days.delete = numeric(n.days);

hours = 0:23;
n.hours = length(hours);
hours.file = hours.insert = hours.delete = numeric(n.hours);

for(i in 1:total.commits)
  {
  my.hour = log.df$day.hour[i];
    idx.hour = which(my.hour == hours)[1];
  my.week = log.df$year.week[i];
    idx.week = which(my.week == weeks)[1];
  my.day = log.df$day.week[i];
    idx.day = which(my.day == days)[1];

  n.file = log.df$file[i];
  n.insert = log.df$insert[i];
  n.delete = log.df$delete[i];

  hours.file[idx.hour]   = hours.file[idx.hour]   + n.file;
```

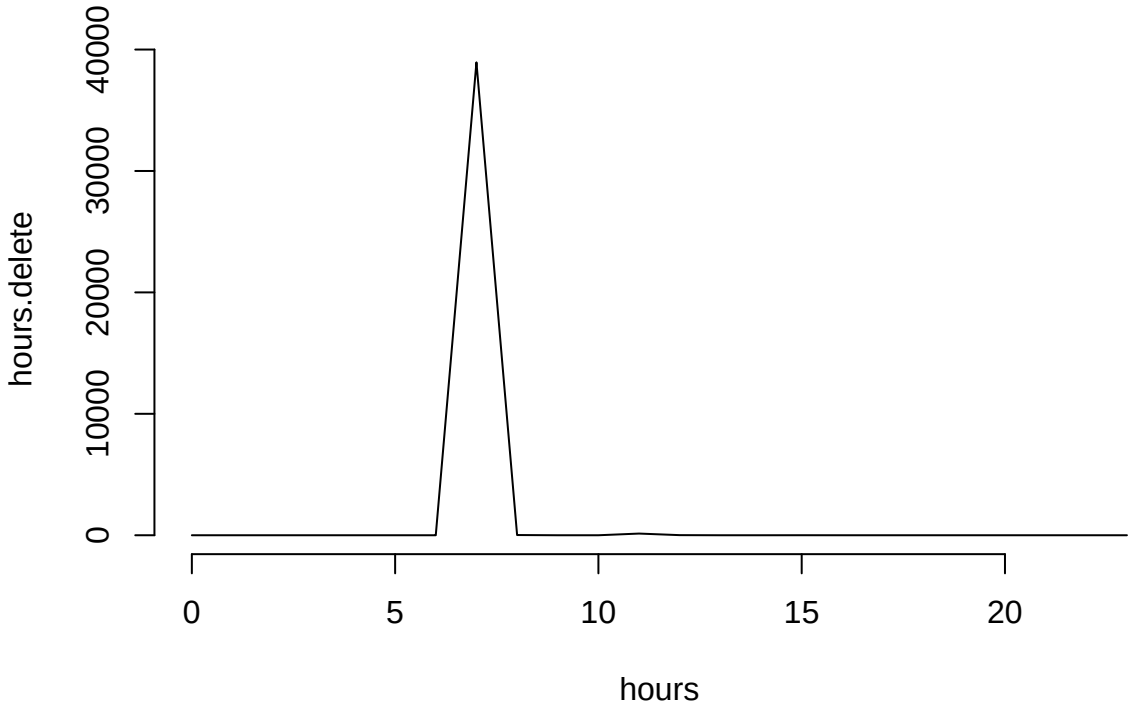
```
  hours.insert[idx.hour] = hours.insert[idx.hour] + n.insert;
  hours.delete[idx.hour] = hours.delete[idx.hour] + n.delete;

  days.file[idx.day]   = days.file[idx.day]   + n.file;
  days.insert[idx.day] = days.insert[idx.day] + n.insert;
  days.delete[idx.day] = days.delete[idx.day] + n.delete;

  weeks.file[idx.week]   = weeks.file[idx.week]   + n.file;
  weeks.insert[idx.week] = weeks.insert[idx.week] + n.insert;
  weeks.delete[idx.week] = weeks.delete[idx.week] + n.delete;
  }


my.git$groups = list();
my.git$groups$hours = list("hours" = hours, "file" = hours.file, "insert" = hours.insert, "delete" = hou
my.git$groups$days = list("days" = days, "file" = days.file, "insert" = days.insert, "delete" = days.del
my.git$groups$weeks = list("weeks" = weeks, "file" = weeks.file, "insert" = weeks.insert, "delete" = wee
my.object$git = my.git;

plot(hours, hours.file,   bty="n", type="l");  # 0 is midnight
```



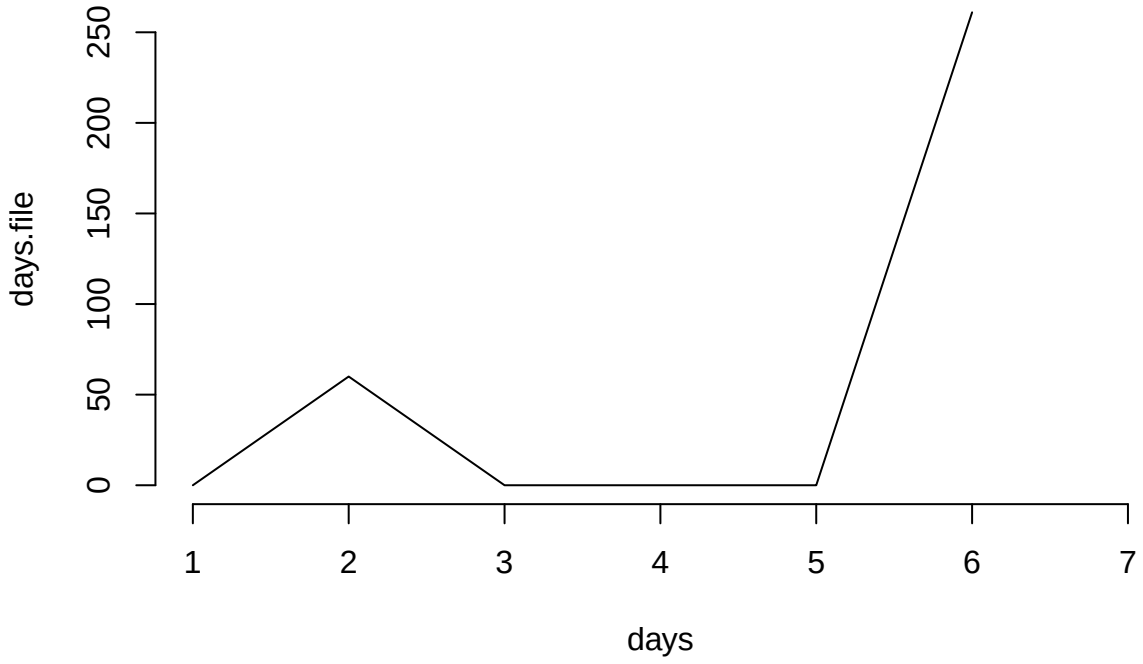
```
plot(hours, hours.insert, bty="n", type="l");
```



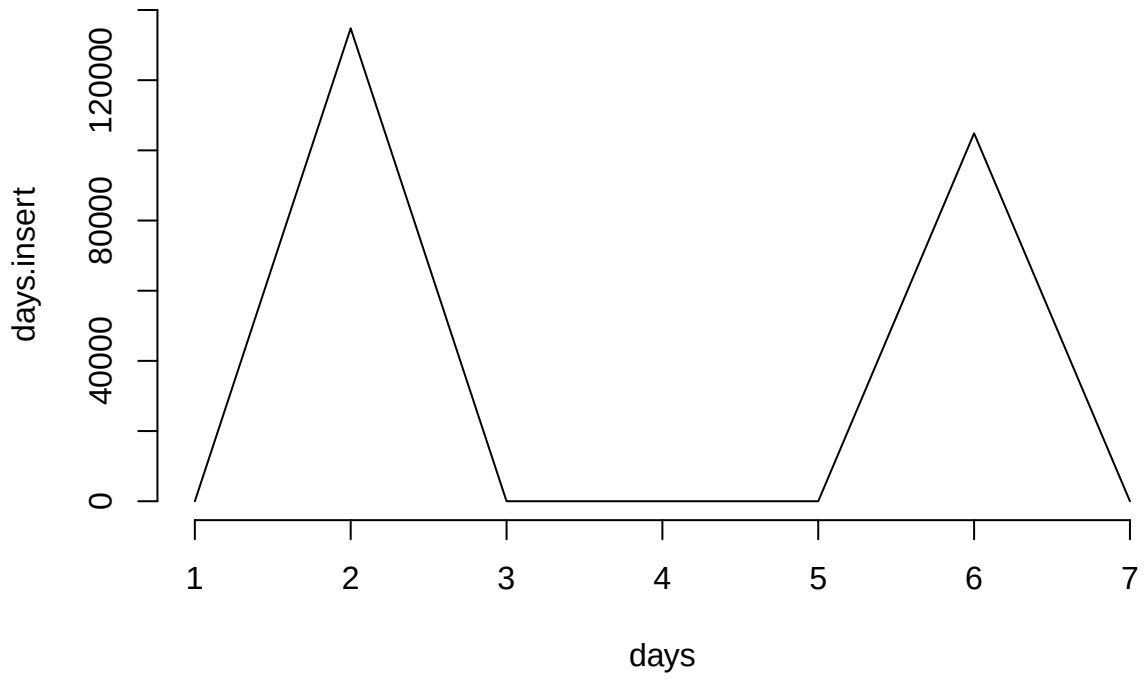

```
plot(hours, hours.delete, bty="n", type="l");
```



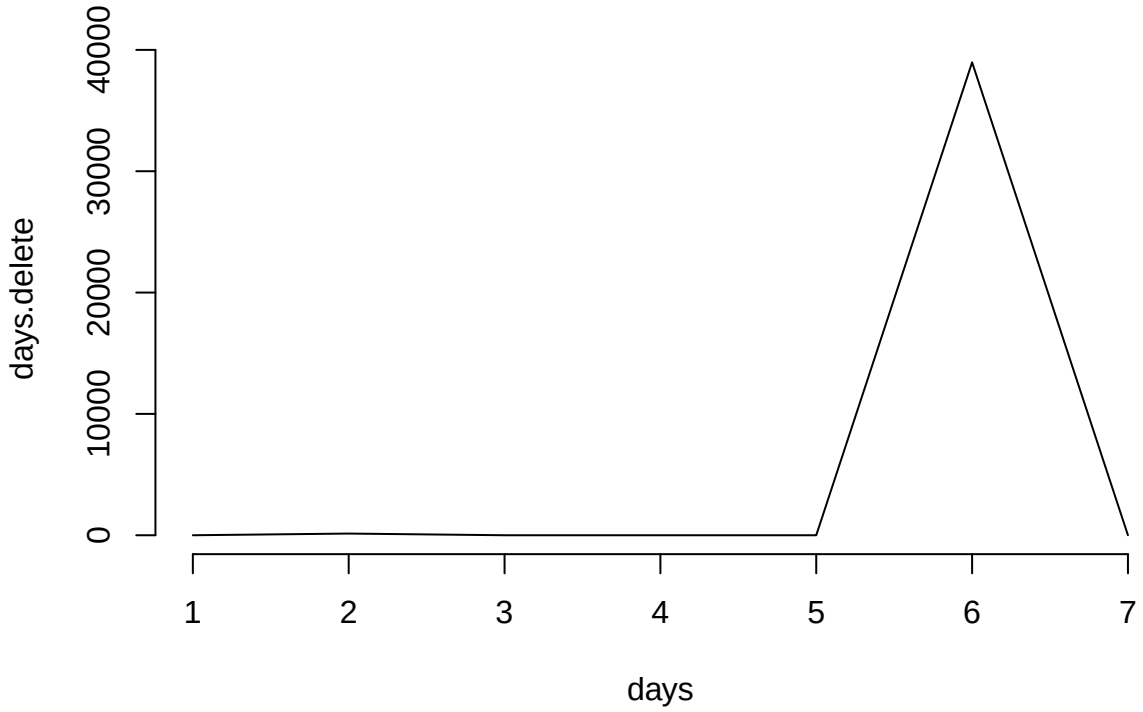

```
plot(days, days.file,   bty="n", type="l");    # 1 is Monday
```



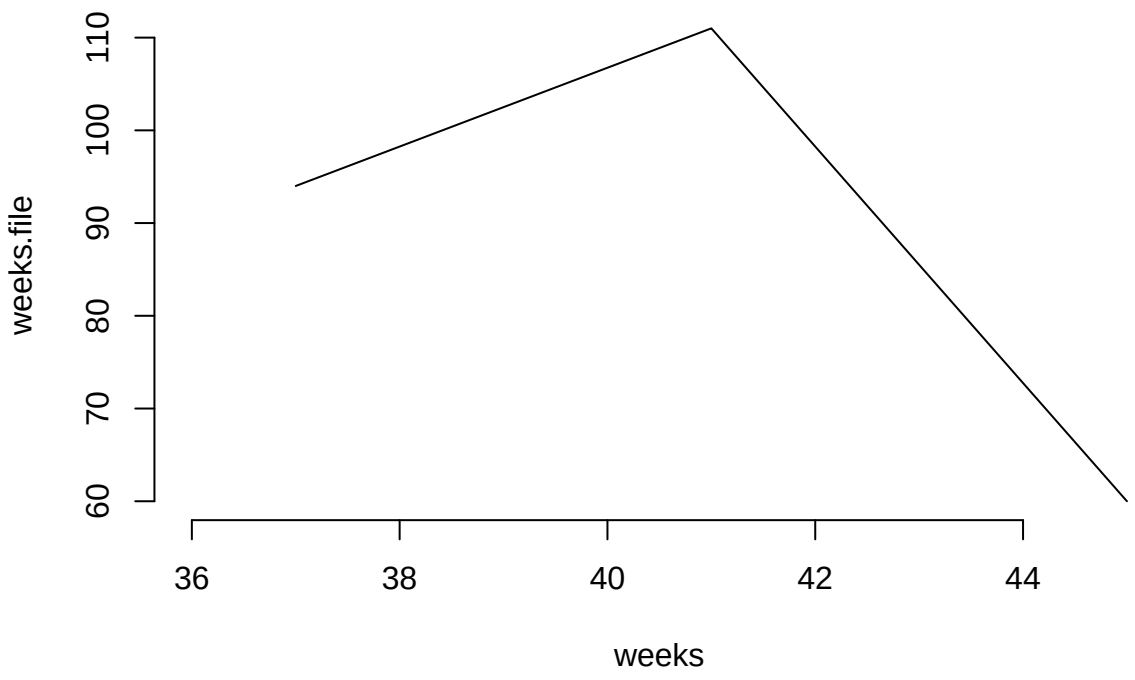

```
plot(days, days.insert, bty="n", type="l");
```



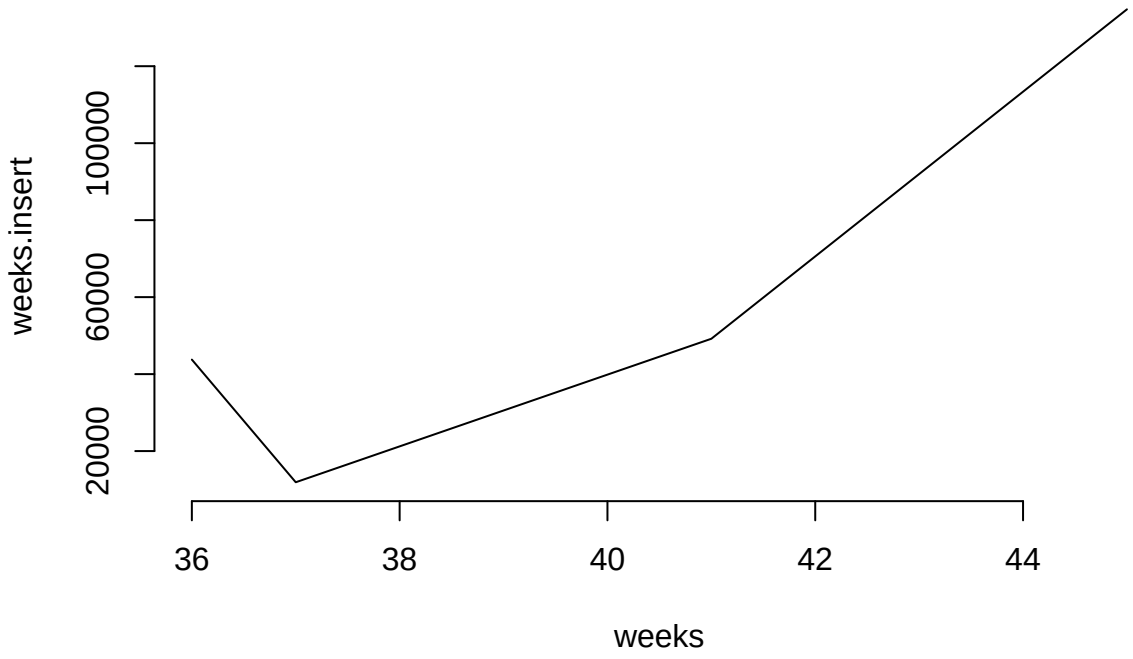

```
plot(days, days.delete, bty="n", type="l");
```



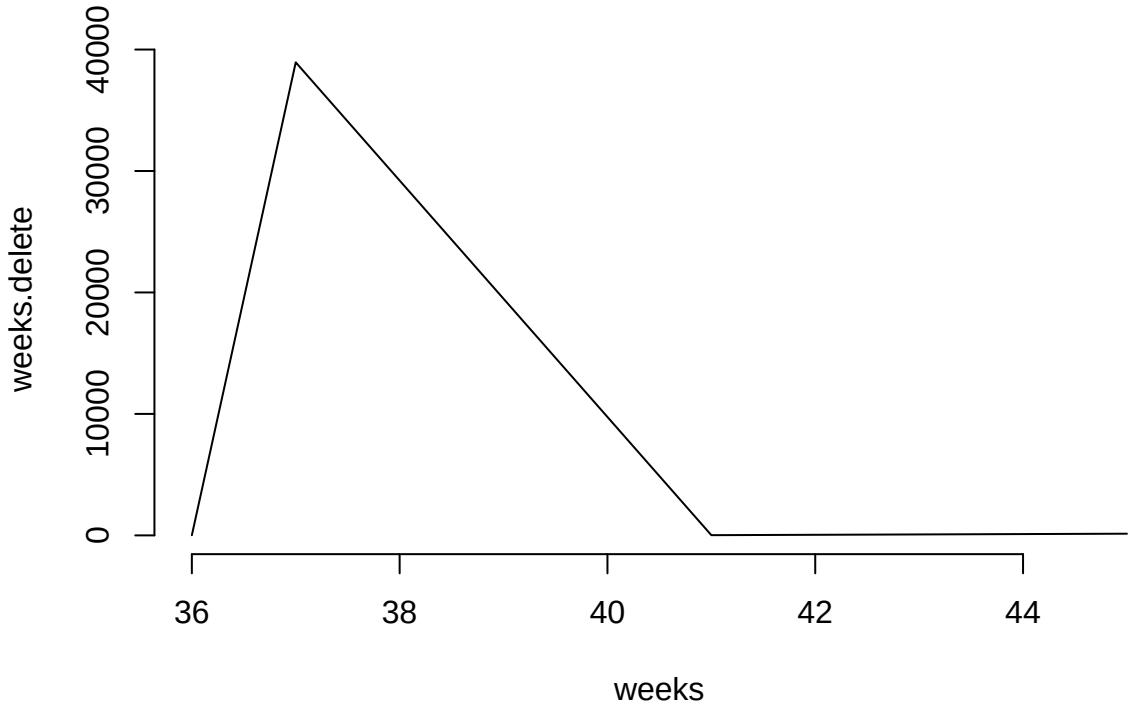

```
plot(weeks, weeks.file,   bty="n", type="l");  # 36 is August 30, 2020
```




```
plot(weeks, weeks.insert, bty="n", type="l");  # https://savvytime.com/week-number
```




```
plot(weeks, weeks.delete, bty="n", type="l");
```




```
################################################
### Feel free to improve on the graphics, if you desire, totally optional ...
```

Conclusive Summary

```
# you need to upload the .rds and .html file after you KNIT ...
saveRDS(my.object, paste0(correct_path,"workspaceAudit.rds") );
```

This audit is worth 50 points. Based on your activity in the workspace, self-assess how many points you merit. Mastery is perfection at 50 points; Developing is about 40 points; Nascent is about 30 points. Justify your decision based on the consistency of your use of the tools and workspace throughout the semester.

Be certain to review the graphs and include self-reflective commentary on your work habits: day of week, hour of day, and so on.