Dementia\_Working\_Final

Isabelle Sanford

1/14/2020

I. HISTORIES a. Sex b. Age c. Region d. Education

1. PRESENT LIFE A. SCORES
   1. General Health
   2. Life Satisfaction
   3. Amount of Mental Activity
   4. Rate of Interaction
   5. Time Pressure B. TIME AMOUNTS
   6. Hours of Personal Care
   7. Hours of Leisure Time
   8. Frequency of Attending Religious Services
2. POTENTIAL ISSUES
3. Age
4. Time Amounts - Binning

This project examined dementia, looking both at the history of people with and without dementia and at their current lives. The former is useful for looking at factors which may impact the likelihood of future dementia, while the latter underscores the difficulties those with dementia face.

The data used are from the University of Michigan’s Panel Study of Income Dynamics, from 2017. The entire family unit was part of the survey, so data is not only self-reported but often also by a spouse or other family member of the household. (Note that this means everyone with a spouse is double-counted while those without are not; a self report is distinct enough from someone else’s perspective that it was judged unhelpful to combine the two in some way. A future report might examine self-reported data vs spouse-reported, or separate them out.)

Dementia was calculated to mean 2 or more of 8 factors, which are regularly used as an assessment of dementia. These are whether in the last several years the person has had a change in: - Making decisions - Amount of interest in activities - Repeating stories or statements over and over - Learning how to use tools (e.g. a TV or phone) - Remembering dates - Trouble with handling money issues - Remembering appointments - Thinking or memory overall

The age range selected was 65 and older, as most dementia patients are elderly and should thus be compared with other elderly people.

I. HISTORIES All of these factors would almost always have been set before the age of 65, and so are considered ‘historical’. The correlations (or lack thereof) with dementia are laid out below.

1. Sex There appears to be no correlation between sex and prevalence of dementia. Out of about 3700 points of data (split evenly between sexes), 20.6% of men and 18.7% of women had dementia, only about 2 percentage points different. With this size of dataset, that difference is likely statistically insignificant.
2. Age As would be expected, age correlates heavily with likelihood of dementia. About 13% of 65-year-olds have dementia, while 45% of 90-year-olds do. (This becomes 90% at 100 years, but there are few enough data points to make that result uncertain at best.)
3. Region If region of the US has an effect, it’s relatively small (14% vs 22% high and low). It’s still worth noting, however, that the Northeast (closely followed by the West) has both the lowest rate of dementia and the smallest number of elderly surveyed (500 as opposed to 1800 in the South).
4. Education Interestingly, education does appear to correlate reasonably strongly with dementia. Those who did not complete high school have by far the highest rate, at 37%. This rate steadily decreases by highest level of education down to just 10% of Masters or equivalent recipients. PhD graduates tick back up to 13% (probably not statistically significant). Higher education in this case may be a proxy for socioeconomic status, which in turn may affect health through living conditions or lack of ability to get treatment. Alternatively, higher education leads to better working conditions: hard labor vs technical / mechanical labor vs non-physical labor. Or perhaps it’s not the state of the working conditions that matters, but the amount to which they require frequent exercise of complex mental abilities.
5. PRESENT LIFE Each of the below variables are about the individual’s life as it presently is.

A. SCORES All of these variables were scored on a scale of 1 to 5, where 5 is “most” of the variable in question and 1 is “least”. Note that this corresponds to 5 being best and 1 worst, except in the case of time pressure.

1. General Health Unsurprisingly, general health status correlates very strongly with dementia, with a correlation coefficient of .96. So all else being statistically equal (injury, etc), dementia probably has a very strong impact on health. (It’s also possible that general physical health being low might make dementia more likely, but this seems an unlikely direction of correlation.)
2. Life Satisfaction Overall life satisfaction is even more highly correlated, but negatively, with a coefficient of -.997. That is, individuals with dementia were much, much more likely to report lower overall life satisfaction than those without. This is fairly intuitive, but the strength of the correlation is admittedly surprising, as is the difference between highest and lowest scores: highest life satisfaction had 12% of individuals having dementia, whereas lowest was 46%.
3. Amount of Mental Activity With a correlation coefficient of -.87, the amount of mental activity an individual engages in is certainly related to dementia (as would make sense). 28% of individuals who reported lowest for mental activity had dementia, as opposed to 11% of individuals who reported highest. An interesting topic of further study (with different data, or in future years if they keep this question on the survey) would be examining the mental activity levels in the past of current dementia patients.
4. Rate of Interaction Somewhat expectedly, interaction correlates negatively with dementia (cor = -.91). This seems likely to be at least in part due to the difficulty for individuals with severe dementia to interact with others, and possibly even remember interacting.
5. Time Pressure How often an individual feels pressed for time does not appear to correlate strongly with dementia: despite its correlation coefficient of .55 (see IIIa), all but one metric fall seemingly randomly within the same six percentage points (15 - 21%), and the last (most pressed for time, at 34%) has significantly less data. A close look at this could prove useful, but it seems more likely to be a statistical error.

B. TIME AMOUNTS Rather than a score between 1 and 5, the variables below are in concrete units representing exactly how often they happen in the individual’s life. Note that values here are more uncertain (see IIIb).

1. Hours of Personal Care (per week) Here, there seems to be a slightly higher proportion of dementia patients among individuals who recorded a low number of hours. Between 0 and 5 hours, 30% of individuals had dementia, while 15 to 20 hours was made up of only 12% dementia patients. This seems reasonable, though may well be more of an implied correlation with age than directly with dementia.
2. Hours of Leisure (per week) Between about 0 and 30 hours per week of leisure time, the proportion of dementia patients seems approximately steady at 20%, i.e. exactly average. Further onward, though, the proportion rises quickly up to a striking 55% at 60 hours per week. This is surprisingly strong, but still understandable: those retired and without dementia are much more likely to spend time on activities they wouldn’t count as leisure, like volunteering, shopping, and so on.
3. Frequency of Attending Religious Services Interestingly, there may be some correlation here with dementia. Individuals who never attended religious services had a 26% portion of dementia, while individuals attending weekly or monthly saw 17% and 12% rates respectively. Yearly visitors were just above weekly at 18%. It’s possible that people with dementia have less motivation or ability to attend services, but there could be some other factor which makes service-goers less likely to have dementia. (That factor doesn’t appear to be religion in general, though, as at a glance the atheist vs religious groups have similar proportions.)
4. POTENTIAL ISSUES
5. Age This analysis may have an underlying factor of age for many of these variables: i.e. some part of the correlation may well because both dementia and X other variable both correlate with age. This seems especially likely with general health and life satisfaction, for instance. This would explain the unusual number of very high correlation coefficients. Given more time or resources\*, a model which takes this into account would be very useful.

\*i.e. if I’d procrastinated less or attended the week on modeling / asked you about it so it wasn’t incredibly confusing

1. Time Amounts - Binning With section IIB, each graph had to choose a bin size that seemed reasonable for the data, in order to make any analysis at all. Given this was due to personal discretion, it may be more likely to be biased in favor of clear trends, and/or there may be clear trends in a bin size different than any tried. Thus the uncertainty in conclusions compared to IIA.

Setup

library(tidyverse)

## -- Attaching packages ---------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.2  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(rvest)

## Loading required package: xml2

##   
## Attaching package: 'rvest'

## The following object is masked from 'package:purrr':  
##   
## pluck

## The following object is masked from 'package:readr':  
##   
## guess\_encoding

library(hexbin)

## Warning: package 'hexbin' was built under R version 3.6.2

library(RColorBrewer)  
library(ggrepel)

## Warning: package 'ggrepel' was built under R version 3.6.2

library(viridisLite)  
library(foreign)

## Warning: package 'foreign' was built under R version 3.6.2

library(tinytex)

## Warning: package 'tinytex' was built under R version 3.6.2

d1 <- read\_csv("Data2.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_double()  
## )

## See spec(...) for full column specifications.

labels <- read\_csv("Labels2.csv")

## Parsed with column specification:  
## cols(  
## Number = col\_character(),  
## Variable = col\_character(),  
## New = col\_character()  
## )

Labeling and Cleaning

labels <- filter(labels, is.na(New) == FALSE)  
names(d1) <- labels$New  
  
  
d2 <- select(d1,-delete1, -delete2, -delete, -delete3, -delete4, -delete5)  
  
  
d3 <- filter(d2, R65plus == 1 | S65plus == 1)

Combining Head & Spouse data

dR <- select(d3, -c(Isex, Sretiredwhen, Sage, Ssex, Shrs\_personal, Shrs\_leisure, Shrs\_volunteer, Sinteraction, Smental\_activity, Stimepressure, Shealth\_general, Swtr\_memloss, Sage\_memloss, Smeds\_memloss, S65plus, Sc\_decisions, Swtr\_hospital, S\_hosp\_nights, S\_hosp\_weeks, Syr\_uni, Smajor\_1, Smajor\_2, Syr\_unigrad, Srelig, Sdenom, Srelig\_svc\_num, Srelig\_svc\_per),-c(Swtr\_hsgrad, Sged\_grade, Shs\_grade, Swtr\_uni, Suni\_grade, Swtr\_unigrad, Shighest\_degree), -c(Sc\_interest, Sc\_story\_repeats, Sc\_learning, Sc\_datemem, Sc\_money, Sc\_apptmem, Sc\_memory, Sc\_decisions), Rsex, Rage) # sometimes an R value  
  
  
dS <- select(d3, FID68, PID68, FID17, PID17, state, wtr\_medical\_bills, medical\_bills, Fwtr\_HI, Fwtr\_noHI, exp\_hosp\_nursing, exp\_presc\_etc, Fmedbills\_amt, Fmedbills\_wtr, exp\_healthbills, Fexp\_hosp\_nurs, Fexp\_presc\_etc, Fexp\_HI, region, weight, FID17\_, seqnum, relation, wtr\_R, F\_wtr\_HI, Isex,   
 retiredwhen = Sretiredwhen,  
 age = Sage,   
 sex = Ssex,   
 hrs\_personal = Shrs\_personal,   
 hrs\_leisure = Shrs\_leisure,   
 hrs\_volunteer = Shrs\_volunteer,   
 interaction = Sinteraction,   
 mental\_activity = Smental\_activity,   
 timepressure = Stimepressure,   
 health\_general = Shealth\_general,   
 wtr\_memloss = Swtr\_memloss,   
 age\_memloss = Sage\_memloss,   
 meds\_memloss = Smeds\_memloss,   
 if65 = S65plus,   
 wtr\_hospital = Swtr\_hospital,   
 hosp\_nights = S\_hosp\_nights,   
 hosp\_weeks = S\_hosp\_weeks,   
 yr\_uni = Syr\_uni,   
 major\_1 = Smajor\_1,   
 major\_2 = Smajor\_2,   
 yr\_unigrad = Syr\_unigrad,   
 relig = Srelig,   
 denom = Sdenom,   
 relig\_svc\_num = Srelig\_svc\_num,   
 relig\_svc\_per = Srelig\_svc\_per,  
 wtr\_hsgrad = Swtr\_hsgrad,   
 ged\_grad = Sged\_grade,   
 hs\_grade = Shs\_grade,   
 wtr\_uni = Swtr\_uni,   
 uni\_grade = Suni\_grade,   
 wtr\_unigrad = Swtr\_unigrad,   
 highest\_degree = Shighest\_degree,  
 c\_interest = Sc\_interest,   
 c\_story\_repeats = Sc\_story\_repeats,   
 c\_learning = Sc\_learning,   
 c\_datemem = Sc\_datemem,   
 c\_money = Sc\_money,   
 c\_apptmem = Sc\_apptmem,   
 c\_memory = Sc\_memory,   
 c\_decisions = Sc\_decisions)  
  
  
dR <- rename(dR, retiredwhen = Rretiredwhen,   
 age = Rage,   
 sex = Rsex,  
 hrs\_personal = Rhrs\_personal,   
 hrs\_leisure = Rhrs\_leisure,   
 hrs\_volunteer = Rhrs\_volunteer,   
 interaction = Rinteraction,   
 mental\_activity = Rmental\_activity,   
 timepressure = Rtimepressure,   
 health\_general = Rhealth\_general,   
 wtr\_memloss = Rwtr\_memloss,   
 age\_memloss = Rage\_memloss,   
 meds\_memloss = Rmeds\_memloss,   
 if65 = R65plus,   
 wtr\_hospital = Rwtr\_hospital,   
 hosp\_nights = R\_hosp\_nights,   
 hosp\_weeks = R\_hosp\_weeks,   
 yr\_uni = Ryr\_uni,   
 major\_1 = Rmajor\_1,   
 major\_2 = Rmajor\_2,   
 yr\_unigrad = Ryr\_unigrad,   
 relig = Rrelig,   
 denom = Rdenom,   
 relig\_svc\_num = Rrelig\_svc\_num,   
 relig\_svc\_per = Rrelig\_svc\_per,   
 wtr\_hsgrad = Rwtr\_hsgrad,   
 ged\_grade = Rged\_grade,   
 hs\_grade = Rhs\_grade,   
 wtr\_uni = Rwtr\_uni,   
 uni\_grade = Runi\_grade,   
 wtr\_unigrad = Rwtr\_unigrad,   
 highest\_degree = Rhighest\_degree,  
 c\_interest = Rc\_interest,   
 c\_story\_repeats = Rc\_story\_repeats,   
 c\_learning = Rc\_learning,   
 c\_datemem = Rc\_datemem,   
 c\_money = Rc\_money,   
 c\_apptmem = Rc\_apptmem,   
 c\_memory = Rc\_memory,   
 c\_decisions = Rc\_decisions)  
  
dcom <- full\_join(dR, dS)

## Joining, by = c("FID68", "PID68", "FID17", "PID17", "state", "age", "sex", "retiredwhen", "hrs\_personal", "hrs\_volunteer", "hrs\_leisure", "interaction", "mental\_activity", "timepressure", "wtr\_medical\_bills", "medical\_bills", "health\_general", "wtr\_memloss", "age\_memloss", "meds\_memloss", "if65", "c\_decisions", "c\_interest", "c\_story\_repeats", "c\_learning", "c\_datemem", "c\_money", "c\_apptmem", "c\_memory", "wtr\_hospital", "hosp\_nights", "hosp\_weeks", "Fwtr\_HI", "Fwtr\_noHI", "exp\_hosp\_nursing", "exp\_presc\_etc", "wtr\_hsgrad", "hs\_grade", "wtr\_uni", "yr\_uni", "uni\_grade", "wtr\_unigrad", "highest\_degree", "major\_1", "major\_2", "yr\_unigrad", "relig", "denom", "relig\_svc\_num", "relig\_svc\_per", "Fmedbills\_wtr", "Fmedbills\_amt", "exp\_healthbills", "Fexp\_hosp\_nurs", "Fexp\_presc\_etc", "Fexp\_HI", "region", "weight", "FID17\_", "seqnum", "relation", "wtr\_R", "F\_wtr\_HI")

Education & Dementia scores

educ <- ifelse(dcom$highest\_degree > 3, 7, ifelse(dcom$highest\_degree == 3, 6, ifelse(dcom$highest\_degree == 2, 5, ifelse(dcom$highest\_degree == 1, 4, ifelse(dcom$wtr\_uni == 1, 3, ifelse(dcom$wtr\_hsgrad < 3, 2, 1))))))  
  
dscore <- ifelse(dcom$c\_decisions == 1, 1, 0) +   
 ifelse(dcom$c\_interest == 1, 1, 0) +   
 ifelse(dcom$c\_story\_repeats == 1, 1, 0) +   
 ifelse(dcom$c\_learning == 1, 1, 0) +   
 ifelse(dcom$c\_datemem == 1, 1, 0) +   
 ifelse(dcom$c\_money == 1, 1, 0) +   
 ifelse(dcom$c\_apptmem == 1, 1, 0) +   
 ifelse(dcom$c\_memory == 1, 1, 0)  
  
  
dcom1 <- dcom %>%   
 select(-c(c\_interest, c\_story\_repeats, c\_learning, c\_datemem, c\_money, c\_apptmem, c\_memory, c\_decisions)) %>%  
 select(-c(wtr\_hsgrad, ged\_grade, hs\_grade, wtr\_uni, uni\_grade, wtr\_unigrad, highest\_degree)) %>%   
 mutate(dscore = dscore, educ = educ)  
  
dwtr <- ifelse(dscore > 1, 1, 0)  
dcom1 <- mutate(dcom1, dwtr = dwtr)

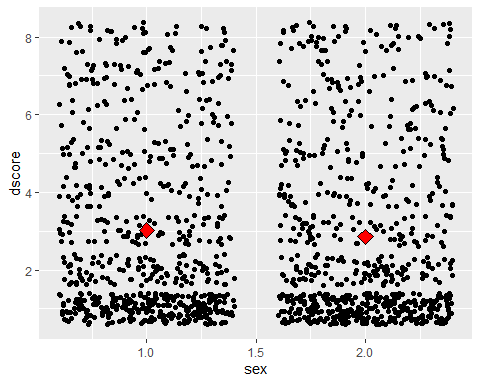
Key: PhD 7 MA 6 Bach 5 Assoc 4 Yr coll 3 HS / GED 2 grade 1-11 > 1

Graph Axis Prep

y\_scale <- scale\_y\_continuous(limits = c(0,.5))

I. HISTORY VARIABLES a. D vs Sex

dcom1 %>%   
 filter(sex > 0, dscore > 0, age > 64) %>%   
 ggplot(aes(sex, dscore)) +   
 geom\_jitter() +  
 stat\_summary(fun.y = mean, geom = "point", shape = 23, size = 4, fill = "red")



dim(filter(dcom1, sex == 1, dwtr == 0, age > 64)) / dim(filter(dcom1, sex == 1, age > 64))

## [1] 0.8021654 1.0000000

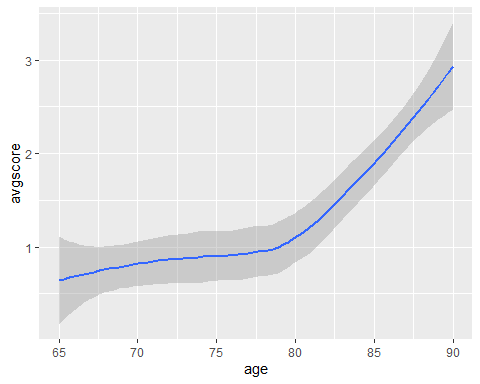
dim(filter(dcom1, sex == 2, dwtr == 0, age > 64)) / dim(filter(dcom1, sex == 2, age > 64))

## [1] 0.8132586 1.0000000

1. Dementia vs Age

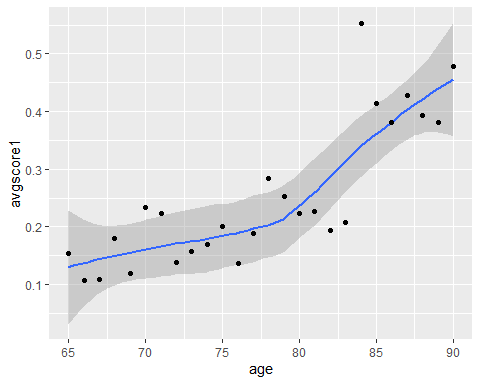
avgscore <- c()  
  
for(i in c(0:25)) {  
 d4 <- filter(dcom1, age == i + 65)  
 avgscore <- append(avgscore, mean(d4$dscore))  
}  
  
age <- c(65:90)  
  
Scoredata <- data.frame(age, avgscore)  
  
ggplot(Scoredata, aes(age, avgscore)) +   
 geom\_smooth()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

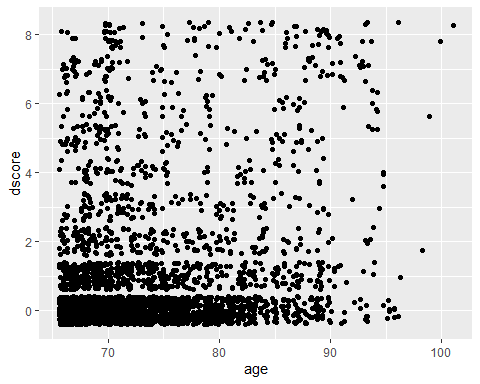


avgscore1 <- c()  
  
for(i in c(0:25)) {  
 d4 <- filter(dcom1, age == i + 65)  
 avg1 <- sum(d4$dwtr) / length(d4$dwtr)  
 avgscore1 <- append(avgscore1, avg1)  
}  
  
  
Scoredata <- data.frame(age, avgscore1)  
  
ggplot(Scoredata, aes(age, avgscore1)) +   
 geom\_smooth() + geom\_point()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



dcom1 %>%   
 filter(age > 65) %>%   
 ggplot(aes(age, dscore)) + geom\_jitter()



cor(Scoredata$age, Scoredata$avgscore1)

## [1] 0.8146907

1. D vs Region (label axis scale!)

f\_rat\_reg <- c()  
m\_rat\_reg <- c()  
all\_rat\_reg <- c()  
  
  
for(i in c(1:4)) {  
 dM <- filter(dcom1, region == i, age > 64, sex == 1)  
 dF <- filter(dcom1, region == i, age > 64, sex == 2)  
 d14 <- filter(dcom1, region == i, age > 64, sex > 0)  
 f\_rat\_reg <- append(f\_rat\_reg, sum(dF$dwtr) / length(dF$region))  
 m\_rat\_reg <- append(m\_rat\_reg, sum(dM$dwtr) / length(dM$region))  
 all\_rat\_reg <- append(all\_rat\_reg, sum(d14$dwtr) / length(d14$region))  
}  
  
  
reg <- c(1:4)  
  
  
mrat\_reg <- data.frame(reg, rat\_reg = m\_rat\_reg, G = "M")  
frat\_reg <- data.frame(reg, rat\_reg = f\_rat\_reg, G = "F")  
allreg <- data.frame(reg, rat\_reg = all\_rat\_reg, G = "A")  
grat <- full\_join(mrat\_reg, frat\_reg)

## Joining, by = c("reg", "rat\_reg", "G")

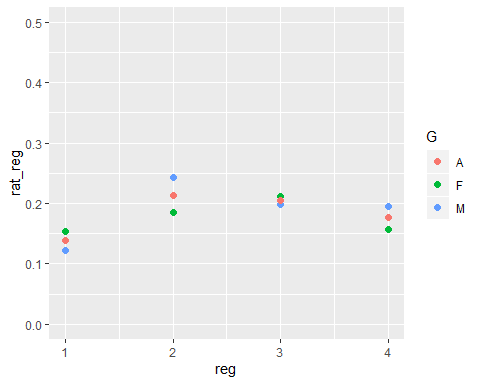
## Warning: Column `G` joining factors with different levels, coercing to  
## character vector

drat <- full\_join(grat, allreg)

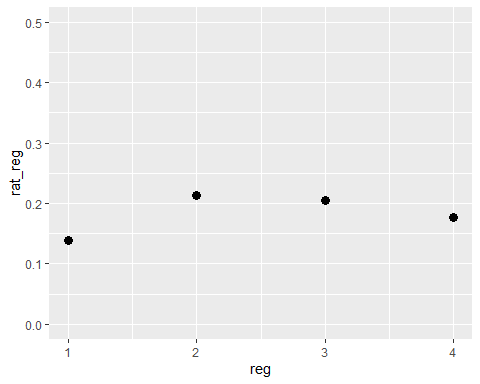
## Joining, by = c("reg", "rat\_reg", "G")

## Warning: Column `G` joining character vector and factor, coercing into  
## character vector

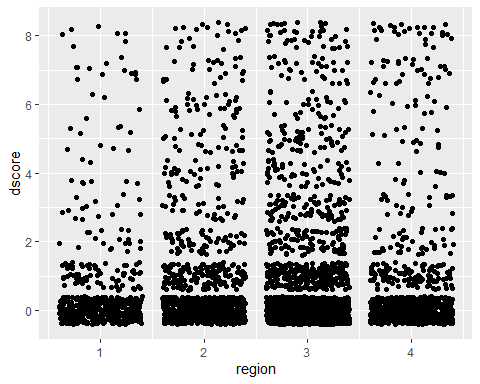
drat$G <- as.factor(drat$G)  
  
ggplot(drat, aes(reg, rat\_reg, color = G)) + geom\_point(size = 2) + y\_scale



ggplot(allreg, aes(reg, rat\_reg)) +   
 geom\_point(size = 3) + y\_scale



dcom1 %>%   
 filter(region < 5, age > 64) %>%   
 ggplot(aes(region, dscore)) + geom\_jitter()



dcom1 %>%   
 filter(age > 64) %>%   
 group\_by(region) %>%   
 summarize(n(), mean(dwtr))

## # A tibble: 6 x 3  
## region `n()` `mean(dwtr)`  
## <dbl> <int> <dbl>  
## 1 1 526 0.139  
## 2 2 986 0.213  
## 3 3 1798 0.205  
## 4 4 829 0.177  
## 5 5 10 0.2   
## 6 6 25 0.04

1. D vs Education (axis scale)

f\_rat\_ed <- c()  
m\_rat\_ed <- c()  
all\_rat\_ed <- c()  
  
  
for(i in c(1:7)) {  
 dM <- filter(dcom1, educ == i, age > 64, sex == 1)  
 dF <- filter(dcom1, educ == i, age > 64, sex == 2)  
 d14 <- filter(dcom1, educ == i, age > 64, sex > 0)  
 f\_rat\_ed <- append(f\_rat\_ed, sum(dF$dwtr) / length(dF$educ))  
 m\_rat\_ed <- append(m\_rat\_ed, sum(dM$dwtr) / length(dM$educ))  
 all\_rat\_ed <- append(all\_rat\_ed, sum(d14$dwtr) / length(d14$educ))  
}  
  
  
ed <- c(1:7)  
  
  
med\_dat <- data.frame(ed, rat\_ed = m\_rat\_ed, G = "M")  
fed\_dat <- data.frame(ed, rat\_ed = f\_rat\_ed, G = "F")  
ded\_dat <- data.frame(ed, rat\_ed = all\_rat\_ed, G = "A")  
ed\_dat <- full\_join(med\_dat, fed\_dat)

## Joining, by = c("ed", "rat\_ed", "G")

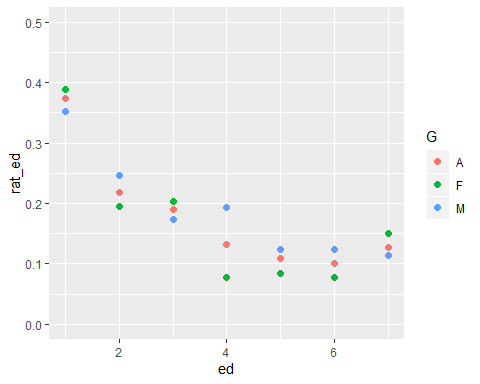
## Warning: Column `G` joining factors with different levels, coercing to  
## character vector

ed\_dat <- full\_join(ed\_dat, ded\_dat)

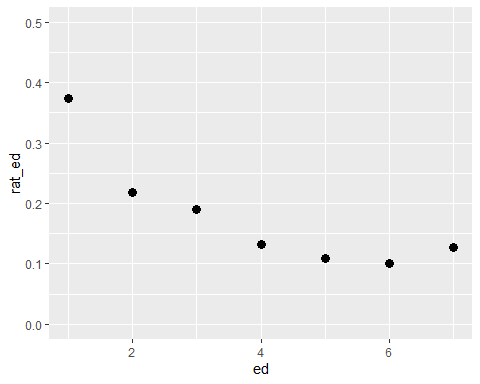
## Joining, by = c("ed", "rat\_ed", "G")

## Warning: Column `G` joining character vector and factor, coercing into  
## character vector

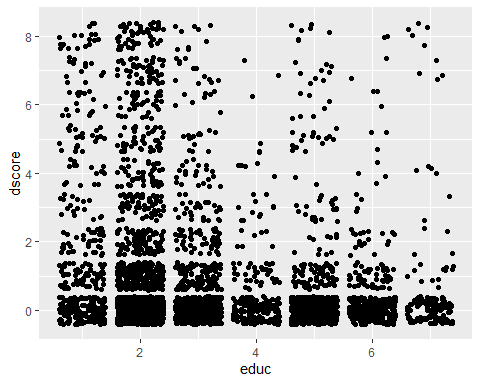
ed\_dat$G <- as.factor(ed\_dat$G)  
  
ggplot(ed\_dat, aes(ed, rat\_ed, color = G)) + geom\_point(size = 2) + y\_scale



ggplot(ded\_dat, aes(ed, rat\_ed)) +   
 geom\_point(size = 3) + y\_scale



dcom1 %>%   
 filter(educ < 8, age > 64) %>%   
 ggplot(aes(educ, dscore)) + geom\_jitter()



dcom1 %>%   
 filter(age > 64, sex == 2) %>%   
 group\_by(educ) %>%   
 summarize(n(), mean(dwtr))

## # A tibble: 7 x 3  
## educ `n()` `mean(dwtr)`  
## <dbl> <int> <dbl>  
## 1 1 229 0.389   
## 2 2 894 0.196   
## 3 3 408 0.203   
## 4 4 116 0.0776  
## 5 5 261 0.0843  
## 6 6 181 0.0773  
## 7 7 53 0.151

1. PRESENT LIFE

A. SCORES a. General Health

dhg <- c()  
fhg <- c()  
mhg <- c()  
hg <- c(1:5)  
  
for(i in c(1:5)) {  
 d6 <- filter(dcom1, health\_general == i, age > 64)  
 dM <- filter(dcom1, health\_general == i, age > 64, sex == 1)  
 dF <- filter(dcom1, health\_general == i, age > 64, sex == 2)  
 dhg <- append(dhg, sum(d6$dwtr) / length(d6$dwtr))  
 fhg <- append(fhg, sum(dF$dwtr) / length(dF$dwtr))  
 mhg <- append(mhg, sum(dM$dwtr) / length(dM$dwtr))  
}  
  
  
dhg\_dat <- data.frame(hg, ahg = dhg, G = "A")  
mhg\_dat <- data.frame(hg, ahg = mhg, G = "M")  
fhg\_dat <- data.frame(hg, ahg = fhg, G = "F")  
hg\_dat <- full\_join(mhg\_dat, full\_join(fhg\_dat, dhg\_dat))

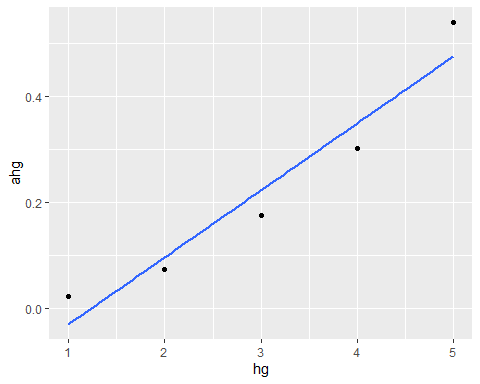
## Joining, by = c("hg", "ahg", "G")

## Warning: Column `G` joining factors with different levels, coercing to  
## character vector

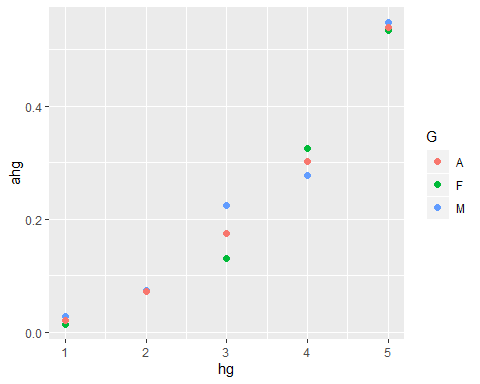
## Joining, by = c("hg", "ahg", "G")

## Warning: Column `G` joining factor and character vector, coercing into  
## character vector

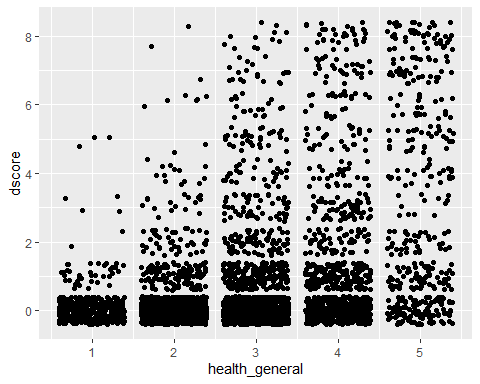
hg\_dat$G <- as.factor(hg\_dat$G)  
  
  
ggplot(dhg\_dat, aes(hg, ahg)) + geom\_point() + geom\_smooth(method = "lm", se = F)



ggplot(hg\_dat, aes(hg, ahg, color = G)) + geom\_point(size = 2)



dcom1 %>%   
 filter(health\_general < 6, age > 64) %>%   
 ggplot(aes(health\_general, dscore)) + geom\_jitter()



cor(hg\_dat$hg, hg\_dat$ahg) # This seems unreasonably high

## [1] 0.9599354

1. D vs Life Satisfaction (points & jitter)

dls <- c()  
fls <- c()  
mls <- c()  
ls <- c(5:1) # because scale is dumb and 1 is best  
  
for(i in c(1:5)) {  
 d6 <- filter(dcom1, life\_satisf == i, age > 64)  
 dM <- filter(dcom1, life\_satisf == i, age > 64, sex == 1)  
 dF <- filter(dcom1, life\_satisf == i, age > 64, sex == 2)  
 dls <- append(dls, sum(d6$dwtr) / length(d6$life\_satisf))  
 fls <- append(fls, sum(dF$dwtr) / length(dF$life\_satisf))  
 mls <- append(mls, sum(dM$dwtr) / length(dM$life\_satisf))  
}  
  
  
dls\_dat <- data.frame(ls, als = dls, G = "A")  
mls\_dat <- data.frame(ls, als = mls, G = "M")  
fls\_dat <- data.frame(ls, als = fls, G = "F")  
ls\_dat <- full\_join(mls\_dat, full\_join(fls\_dat, dls\_dat))

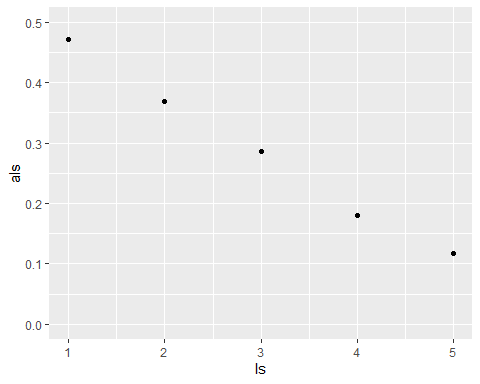
## Joining, by = c("ls", "als", "G")

## Warning: Column `G` joining factors with different levels, coercing to  
## character vector

## Joining, by = c("ls", "als", "G")

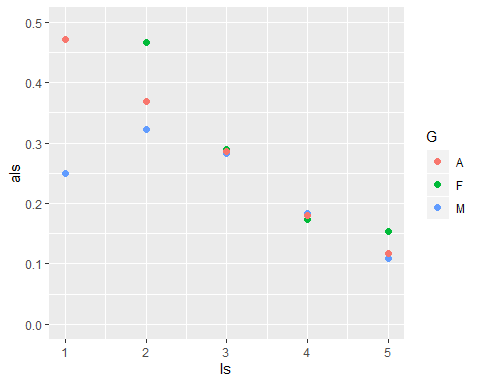
## Warning: Column `G` joining factor and character vector, coercing into  
## character vector

ls\_dat$G <- as.factor(ls\_dat$G)  
  
  
ggplot(dls\_dat, aes(ls, als)) + geom\_point() + y\_scale

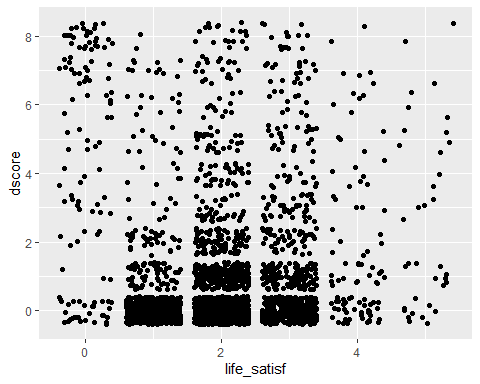


ggplot(ls\_dat, aes(ls, als, color = G)) + geom\_point(size = 2) + y\_scale

## Warning: Removed 1 rows containing missing values (geom\_point).



dcom1 %>%   
 filter(life\_satisf < 6, age > 64) %>%   
 ggplot(aes(life\_satisf, dscore)) + geom\_jitter()



cor(dls\_dat$ls, dls\_dat$als)

## [1] -0.9972192

1. D vs Mental Activity (points)

dma <- c()  
fma <- c()  
mma <- c()  
ma <- c(1:5)  
  
for(i in c(1:5)) {  
 d6 <- filter(dcom1, mental\_activity == i, age > 64)  
 dM <- filter(dcom1, mental\_activity == i, age > 64, sex == 1)  
 dF <- filter(dcom1, mental\_activity == i, age > 64, sex == 2)  
 dma <- append(dma, sum(d6$dwtr) / length(d6$mental\_activity))  
 fma <- append(fma, sum(dF$dwtr) / length(dF$mental\_activity))  
 mma <- append(mma, sum(dM$dwtr) / length(dM$mental\_activity))  
}  
  
  
dma\_dat <- data.frame(ma, ama = dma, G = "A")  
mma\_dat <- data.frame(ma, ama = mma, G = "M")  
fma\_dat <- data.frame(ma, ama = fma, G = "F")  
ma\_dat <- full\_join(mma\_dat, full\_join(fma\_dat, dma\_dat))

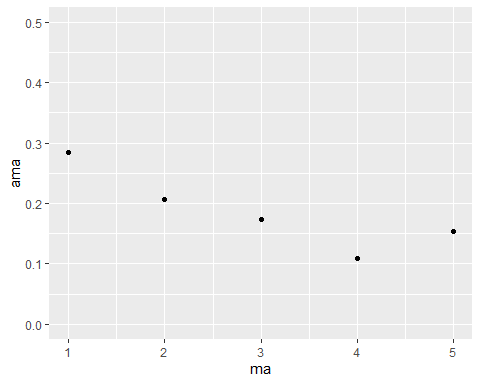
## Joining, by = c("ma", "ama", "G")

## Warning: Column `G` joining factors with different levels, coercing to  
## character vector

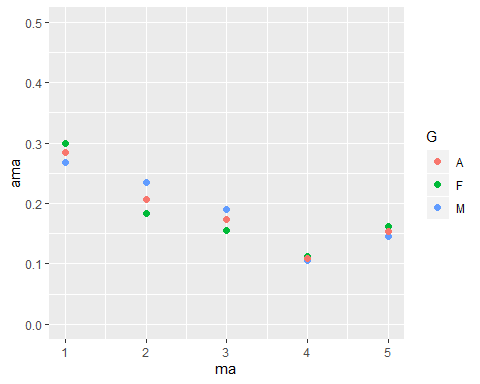
## Joining, by = c("ma", "ama", "G")

## Warning: Column `G` joining factor and character vector, coercing into  
## character vector

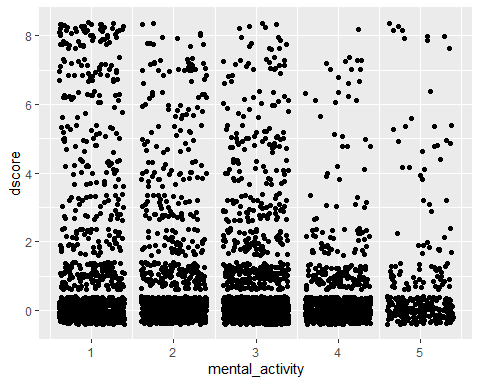
ma\_dat$G <- as.factor(ma\_dat$G)  
  
  
ggplot(dma\_dat, aes(ma, ama)) + geom\_point() + y\_scale



ggplot(ma\_dat, aes(ma, ama, color = G)) + geom\_point(size = 2) + y\_scale



dcom1 %>%   
 filter(mental\_activity < 6, age > 64) %>%   
 ggplot(aes(mental\_activity, dscore)) + geom\_jitter()



cor(dma\_dat$ma, dma\_dat$ama)

## [1] -0.8652993

1. D vs Interaction (points)

dint <- c()  
fint <- c()  
mint <- c()  
int <- c(1:5)  
  
for(i in c(1:5)) {  
 d6 <- filter(dcom1, interaction == i, age > 64)  
 dM <- filter(dcom1, interaction == i, age > 64, sex == 1)  
 dF <- filter(dcom1, interaction == i, age > 64, sex == 2)  
 dint <- append(dint, sum(d6$dwtr) / length(d6$interaction))  
 fint <- append(fint, sum(dF$dwtr) / length(dF$interaction))  
 mint <- append(mint, sum(dM$dwtr) / length(dM$interaction))  
}  
  
  
dint\_dat <- data.frame(int, aint = dint, G = "A")  
mint\_dat <- data.frame(int, aint = mint, G = "M")  
fint\_dat <- data.frame(int, aint = fint, G = "F")  
int\_dat <- full\_join(mint\_dat, full\_join(fint\_dat, dint\_dat))

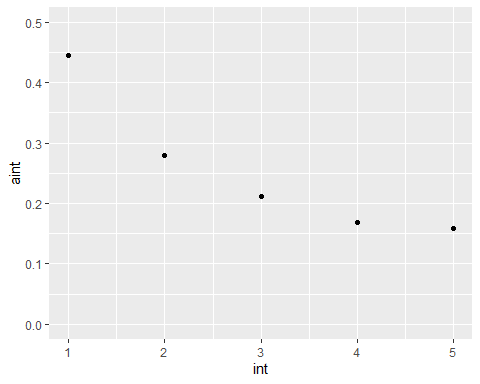
## Joining, by = c("int", "aint", "G")

## Warning: Column `G` joining factors with different levels, coercing to  
## character vector

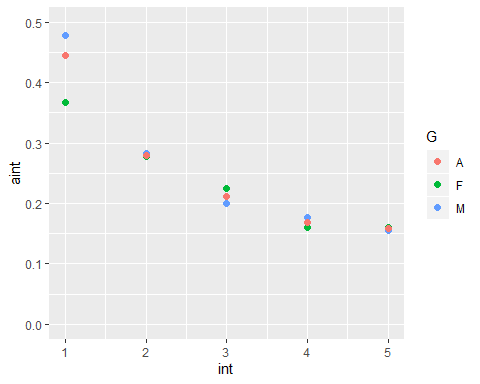
## Joining, by = c("int", "aint", "G")

## Warning: Column `G` joining factor and character vector, coercing into  
## character vector

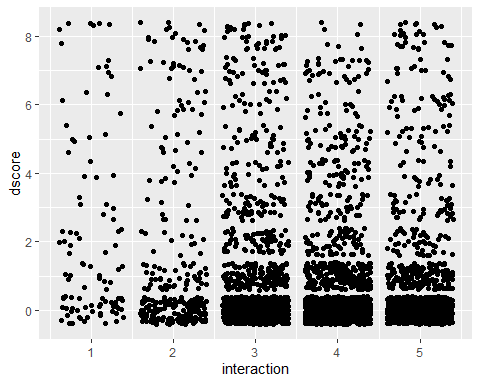
int\_dat$G <- as.factor(int\_dat$G)  
  
  
ggplot(dint\_dat, aes(int, aint)) + geom\_point() + y\_scale



ggplot(int\_dat, aes(int, aint, color = G)) + geom\_point(size = 2) + y\_scale



dcom1 %>%   
 filter(interaction < 6, age > 64) %>%   
 ggplot(aes(interaction, dscore)) + geom\_jitter()



cor(dint\_dat$aint, dint\_dat$int)

## [1] -0.9198645

(bcd) Combination of previous 3

dma\_dat <- rename(dma\_dat, vec = ma)  
dls\_dat <- rename(dls\_dat, vec = ls)  
dint\_dat <- rename(dint\_dat, vec = int)  
  
dat3 <- full\_join(dma\_dat, full\_join(dls\_dat, dint\_dat))

## Joining, by = c("vec", "G")  
## Joining, by = c("vec", "G")

dat3

## vec ama G als aint  
## 1 1 0.2841163 A 0.4722222 0.4444444  
## 2 2 0.2069328 A 0.3695652 0.2802768  
## 3 3 0.1733840 A 0.2857143 0.2113402  
## 4 4 0.1087866 A 0.1806303 0.1680556  
## 5 5 0.1535714 A 0.1177347 0.1587575

avg <- c()  
  
for(i in c(1:5)) {  
# filter(dat3, vec == i)  
 vec1 <- (dat3$ama[i] + dat3$als[i] + dat3$aint[i]) / 3  
 avg <- append(avg, vec1)  
}  
  
avg

## [1] 0.4002610 0.2855916 0.2234795 0.1524908 0.1433546

ggplot(dat3, aes(vec)) +   
 geom\_point(aes(y = ama), color = "purple") +   
 geom\_point(aes(y = als), color = "red") +  
 geom\_point(aes(y = aint), color = "green") +  
 geom\_smooth(aes(y = avg)) +  
 y\_scale

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : pseudoinverse used at 0.98

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : neighborhood radius 2.02

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : There are other near singularities as well. 4.0804

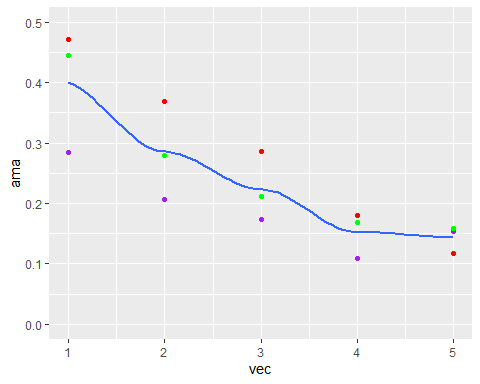
## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : span too small.  
## fewer data values than degrees of freedom.

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used  
## at 0.98

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius  
## 2.02

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal  
## condition number 0

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : There are other  
## near singularities as well. 4.0804



1. D vs Time Pressure (points & jitter)

dtp <- c()  
ftp <- c()  
mtp <- c()  
tp <- c(1:5)  
  
for(i in c(1:5)) {  
 d6 <- filter(dcom1, timepressure == i, age > 64)  
 dM <- filter(dcom1, timepressure == i, age > 64, sex == 1)  
 dF <- filter(dcom1, timepressure == i, age > 64, sex == 2)  
 dtp <- append(dtp, sum(d6$dwtr) / length(d6$timepressure))  
 ftp <- append(ftp, sum(dF$dwtr) / length(dF$timepressure))  
 mtp <- append(mtp, sum(dM$dwtr) / length(dM$timepressure))  
}  
  
  
dtp\_dat <- data.frame(tp, atp = dtp, G = "A")  
mtp\_dat <- data.frame(tp, atp = mtp, G = "M")  
ftp\_dat <- data.frame(tp, atp = ftp, G = "F")  
tp\_dat <- full\_join(mtp\_dat, full\_join(ftp\_dat, dtp\_dat))

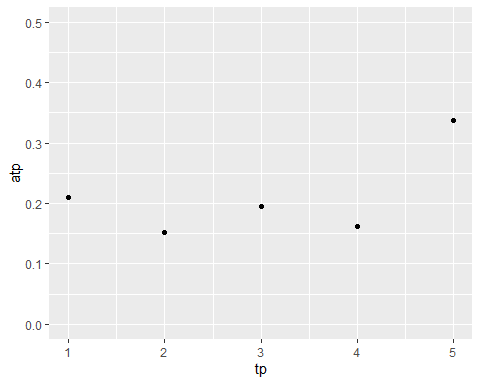
## Joining, by = c("tp", "atp", "G")

## Warning: Column `G` joining factors with different levels, coercing to  
## character vector

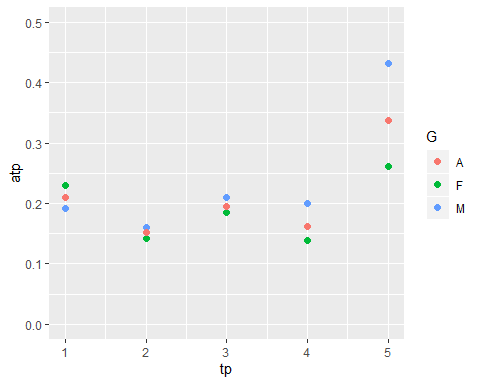
## Joining, by = c("tp", "atp", "G")

## Warning: Column `G` joining factor and character vector, coercing into  
## character vector

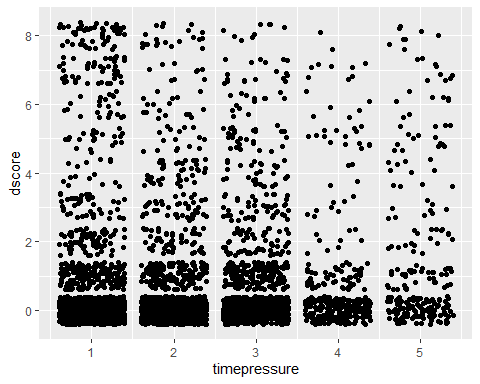
tp\_dat$G <- as.factor(tp\_dat$G)  
  
  
ggplot(dtp\_dat, aes(tp, atp)) + geom\_point() + y\_scale



ggplot(tp\_dat, aes(tp, atp, color = G)) + geom\_point(size = 2) + y\_scale



dcom1 %>%   
 filter(timepressure < 6, age > 64) %>%   
 ggplot(aes(timepressure, dscore)) + geom\_jitter()



cor(dtp\_dat$atp, dtp\_dat$tp)

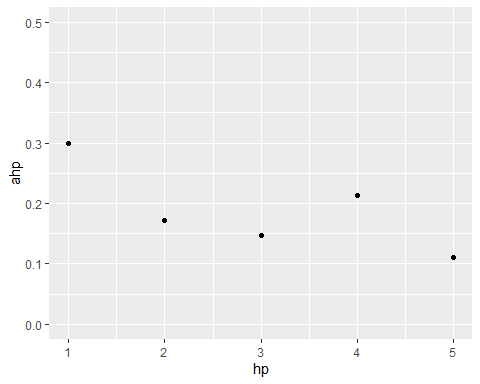
## [1] 0.5598491

dcom1 %>%   
 filter(timepressure < 6, age > 64) %>%   
 group\_by(timepressure) %>%   
 summarise(n())

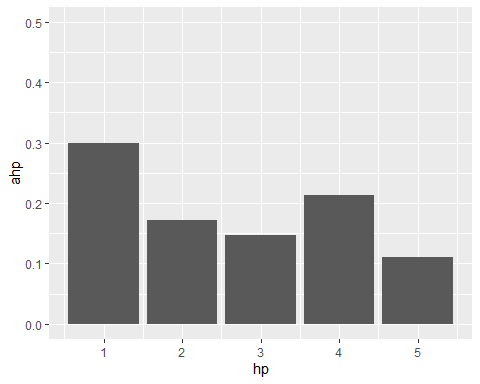
## # A tibble: 5 x 2  
## timepressure `n()`  
## <dbl> <int>  
## 1 1 1255  
## 2 2 1292  
## 3 3 1045  
## 4 4 322  
## 5 5 234

B. TIME AMOUNTS a. D vs Personal Hours (points & jitter)

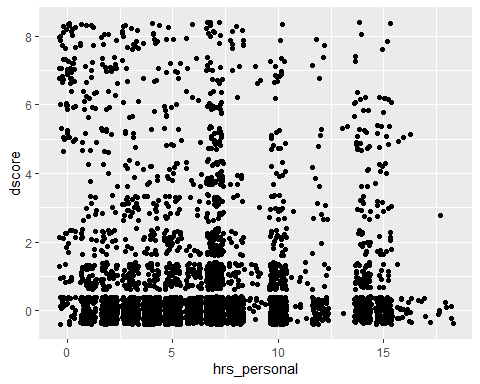
dhp <- c()  
fhp <- c()  
mhp <- c()  
hp <- c(1:5)  
  
for(i in c(1:5)) {  
 d6 <- filter(dcom1, hrs\_personal < i\*4, hrs\_personal >= (i-1)\*4, age > 64)  
# dM <- filter(dcom1, hrs\_personal == i, age > 64, sex == 1)  
# dF <- filter(dcom1, hrs\_personal == i, age > 64, sex == 2)  
 dhp <- append(dhp, sum(d6$dwtr) / length(d6$hrs\_personal))  
# fhp <- append(fhp, sum(dF$dwtr) / length(dF$hrs\_personal))  
# mhp <- append(mhp, sum(dM$dwtr) / length(dM$hrs\_personal))  
}  
  
  
dhp\_dat <- data.frame(hp, ahp = dhp, G = "A")  
#mhp\_dat <- data.frame(hp, ahp = mhp, G = "M")  
#fhp\_dat <- data.frame(hp, ahp = fhp, G = "F")  
#hp\_dat <- full\_join(mhp\_dat, full\_join(fhp\_dat, dhp\_dat))  
#hp\_dat$G <- as.factor(hp\_dat$G)  
  
  
ggplot(dhp\_dat, aes(hp, ahp)) + geom\_point() + y\_scale



ggplot(dhp\_dat, aes(hp, ahp)) + geom\_col() + y\_scale



#ggplot(hp\_dat, aes(hp, ahp, color = G)) + geom\_point(size = 2) + geom\_smooth(se = FALSE, method = "lm") + y\_scale  
  
dcom1 %>%   
 filter(hrs\_personal < 20, age > 64) %>%   
 ggplot(aes(hrs\_personal, dscore)) + geom\_jitter()



cor(dhp\_dat$hp, dhp\_dat$ahp)

## [1] -0.7313883

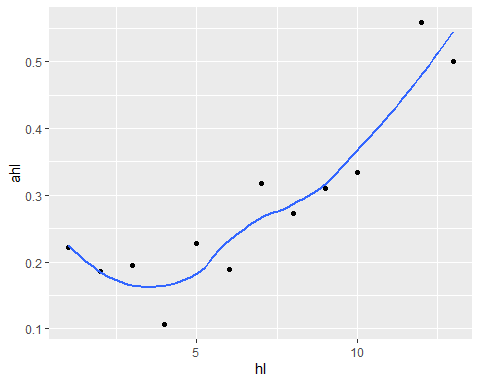
1. D vs Leisure Hours (pt, jitter, smooth needs work)

dhl <- c()  
fhl <- c()  
mhl <- c()  
hl <- c(1:13)  
  
for(i in c(1:13)) {  
 d6 <- filter(dcom1, hrs\_leisure < i\*5, hrs\_leisure > (i-1)\*5, age > 64)  
 #dM <- filter(dcom1, hrs\_leisure == i, age > 64, sex == 1)  
 #dF <- filter(dcom1, hrs\_leisure == i, age > 64, sex == 2)  
 dhl <- append(dhl, sum(d6$dwtr) / length(d6$hrs\_leisure))  
 #fhl <- append(fhl, sum(dF$dwtr) / length(dF$hrs\_leisure))  
 #mhl <- append(mhl, sum(dM$dwtr) / length(dM$hrs\_leisure))  
}  
  
  
dhl\_dat <- data.frame(hl, ahl = dhl, G = "A")  
#mhl\_dat <- data.frame(hl, ahl = mhl, G = "M")  
#fhl\_dat <- data.frame(hl, ahl = fhl, G = "F")  
#hl\_dat <- full\_join(mhl\_dat, full\_join(fhl\_dat, dhl\_dat))  
#hl\_dat$G <- as.factor(hl\_dat$G)  
  
  
ggplot(dhl\_dat, aes(hl, ahl)) + geom\_point() + geom\_smooth(se = F) #+ y\_scale

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (stat\_smooth).

## Warning: Removed 1 rows containing missing values (geom\_point).

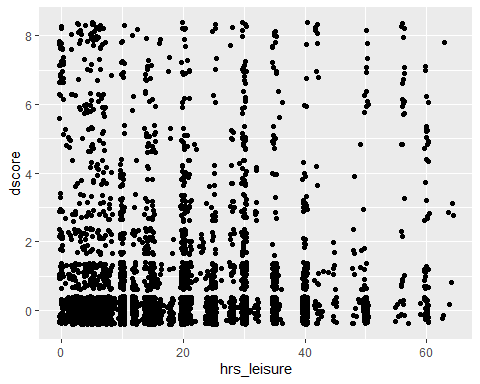


ggplot(dhl\_dat, aes(hl, ahl)) + geom\_col() #+ y\_scale

## Warning: Removed 1 rows containing missing values (position\_stack).



#ggplot(hl\_dat, aes(hl, ahl, color = G)) + geom\_point(size = 2) + y\_scale  
  
dcom1 %>%   
 filter(hrs\_leisure < 65, age > 64) %>%   
 ggplot(aes(hrs\_leisure, dscore)) + geom\_jitter()



1. D vs # Services (pts & jitter) (CHECK KEY)

dsvp <- c()  
fsvp <- c()  
msvp <- c()  
svp <- c(0,3,5,6)  
  
for(i in c(0,3,5,6)) {  
 d6 <- filter(dcom1, relig\_svc\_per == i, age > 64)  
 dM <- filter(dcom1, relig\_svc\_per == i, age > 64, sex == 1)  
 dF <- filter(dcom1, relig\_svc\_per == i, age > 64, sex == 2)  
 dsvp <- append(dsvp, sum(d6$dwtr) / length(d6$relig\_svc\_per))  
 fsvp <- append(fsvp, sum(dF$dwtr) / length(dF$relig\_svc\_per))  
 msvp <- append(msvp, sum(dM$dwtr) / length(dM$relig\_svc\_per))  
}  
  
  
dsvp\_dat <- data.frame(svp, asvp = dsvp, G = "A")  
msvp\_dat <- data.frame(svp, asvp = msvp, G = "M")  
fsvp\_dat <- data.frame(svp, asvp = fsvp, G = "F")  
svp\_dat <- full\_join(msvp\_dat, full\_join(fsvp\_dat, dsvp\_dat))

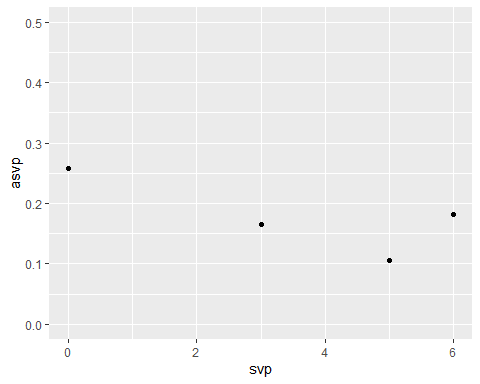
## Joining, by = c("svp", "asvp", "G")

## Warning: Column `G` joining factors with different levels, coercing to  
## character vector

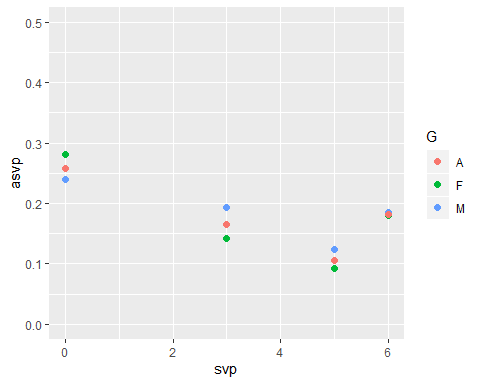
## Joining, by = c("svp", "asvp", "G")

## Warning: Column `G` joining factor and character vector, coercing into  
## character vector

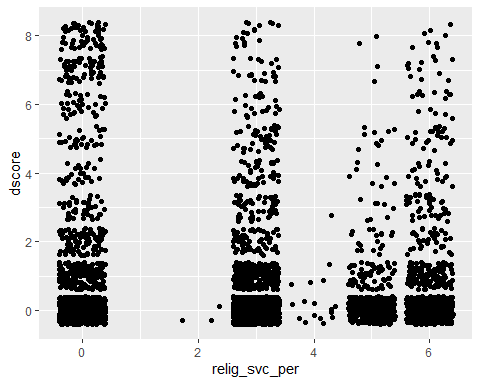
svp\_dat$G <- as.factor(svp\_dat$G)  
  
  
ggplot(dsvp\_dat, aes(svp, asvp)) + geom\_point() + y\_scale



ggplot(svp\_dat, aes(svp, asvp, color = G)) + geom\_point(size = 2) + y\_scale



dcom1 %>%   
 filter(relig\_svc\_per < 7, age > 64) %>%   
 ggplot(aes(relig\_svc\_per, dscore)) + geom\_jitter()



dcom1 %>%   
 mutate(relyn = ifelse(relig == 0, 0, 1)) %>%   
 group\_by(relyn) %>%   
 filter(age < 64) %>%   
 summarise(n(), sum(dwtr))

## # A tibble: 2 x 3  
## relyn `n()` `sum(dwtr)`  
## <dbl> <int> <dbl>  
## 1 0 1069 0  
## 2 1 616 0

dcom1 %>%   
 filter(age > 64) %>%   
 group\_by(relig) %>%   
 summarise(sum(dwtr) / n())

## # A tibble: 8 x 2  
## relig `sum(dwtr)/n()`  
## <dbl> <dbl>  
## 1 0 0.152   
## 2 1 0.178   
## 3 2 0.219   
## 4 8 0.196   
## 5 10 0.137   
## 6 13 0.0714  
## 7 97 0.267   
## 8 99 0.301

# yes I should have done everything this way but I only realized that just now oops

Whether Hospitalized

dwh <- c()  
fwh <- c()  
mwh <- c()  
wh <- c(1,5)  
  
for(i in c(1,5)) {  
 d6 <- filter(dcom1, wtr\_hospital == i, age > 64)  
 dM <- filter(dcom1, wtr\_hospital == i, age > 64, sex == 1)  
 dF <- filter(dcom1, wtr\_hospital == i, age > 64, sex == 2)  
 dwh <- append(dwh, sum(d6$dwtr) / length(d6$wtr\_hospital))  
 fwh <- append(fwh, sum(dF$dwtr) / length(dF$wtr\_hospital))  
 mwh <- append(mwh, sum(dM$dwtr) / length(dM$wtr\_hospital))  
}  
  
  
dwh\_dat <- data.frame(wh, awh = dwh, G = "A")  
mwh\_dat <- data.frame(wh, awh = mwh, G = "M")  
fwh\_dat <- data.frame(wh, awh = fwh, G = "F")  
wh\_dat <- full\_join(mwh\_dat, full\_join(fwh\_dat, dwh\_dat))

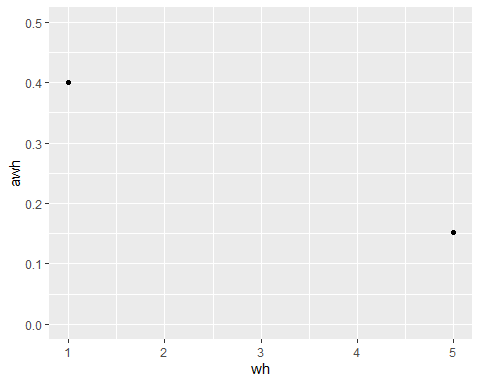
## Joining, by = c("wh", "awh", "G")

## Warning: Column `G` joining factors with different levels, coercing to  
## character vector

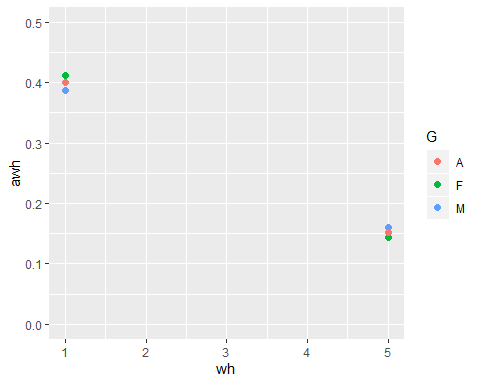
## Joining, by = c("wh", "awh", "G")

## Warning: Column `G` joining factor and character vector, coercing into  
## character vector

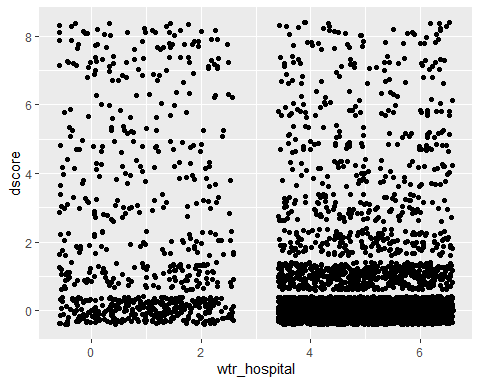
wh\_dat$G <- as.factor(wh\_dat$G)  
  
  
ggplot(dwh\_dat, aes(wh, awh)) + geom\_point() + y\_scale



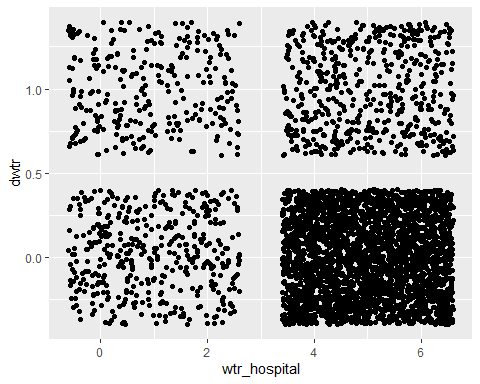
ggplot(wh\_dat, aes(wh, awh, color = G)) + geom\_point(size = 2) + y\_scale



dcom1 %>%   
 filter(wtr\_hospital < 6, age > 64) %>%   
 ggplot(aes(wtr\_hospital, dscore)) + geom\_jitter()



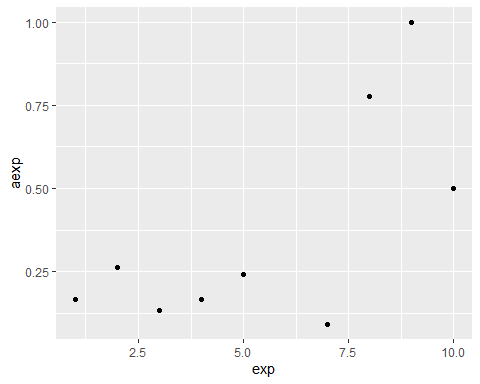
dcom1 %>%   
 filter(age > 64, wtr\_hospital < 6) %>%   
 ggplot(aes(wtr\_hospital, dwtr)) + geom\_jitter()



Prescription & Other Expenses

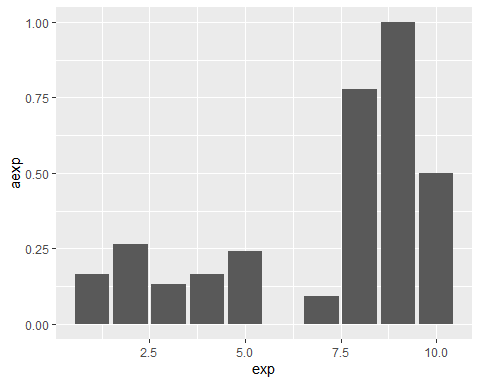
dexp <- c()  
fexp <- c()  
mexp <- c()  
exp <- c(1:10)  
  
for(i in c(1:10)) {  
 d6 <- filter(dcom1, exp\_presc\_etc <= i\*1000, exp\_presc\_etc > (i - 1)\*1000, age > 64)  
 dexp <- append(dexp, sum(d6$dwtr) / length(d6$exp\_presc\_etc))  
}  
  
  
dexp\_dat <- data.frame(exp, aexp = dexp, G = "A")  
  
  
ggplot(dexp\_dat, aes(exp, aexp)) + geom\_point()

## Warning: Removed 1 rows containing missing values (geom\_point).



ggplot(dexp\_dat, aes(exp, aexp)) + geom\_col()

## Warning: Removed 1 rows containing missing values (position\_stack).



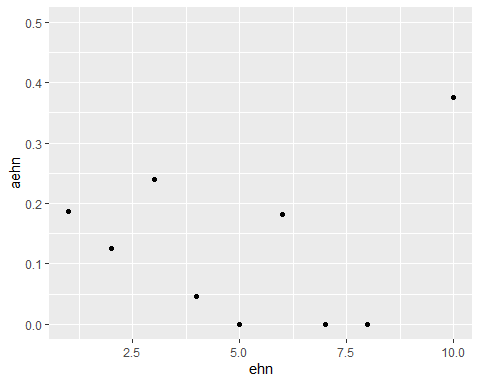
#ggplot(exp\_dat, aes(exp, aexp, color = G)) + geom\_point(size = 2)  
  
dcom1 %>%   
 filter(exp\_presc\_etc < 10000, age > 64) %>%   
 ggplot(aes(exp\_presc\_etc, dwtr)) + geom\_jitter()



Hospital & Nursing Expenses

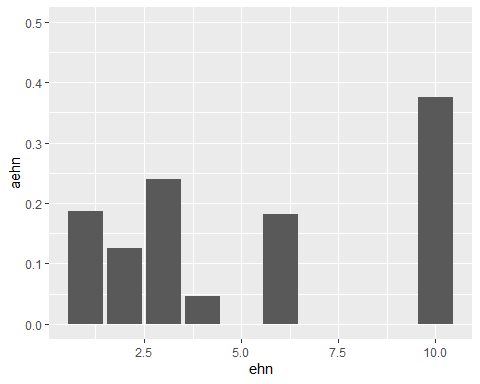
dehn <- c()  
fehn <- c()  
mehn <- c()  
ehn <- c(1:10)  
  
for(i in c(1:10)) {  
 d6 <- filter(dcom1,   
 exp\_hosp\_nursing <= i\*1000,   
 exp\_hosp\_nursing > (i - 1)\*1000,   
 age > 64,  
 wtr\_hospital == 5)  
 dehn <- append(dehn, sum(d6$dwtr) / length(d6$exp\_hosp\_nursing))  
}  
  
  
dehn\_dat <- data.frame(ehn, aehn = dehn, G = "A")  
  
  
ggplot(dehn\_dat, aes(ehn, aehn)) + geom\_point() + y\_scale

## Warning: Removed 1 rows containing missing values (geom\_point).

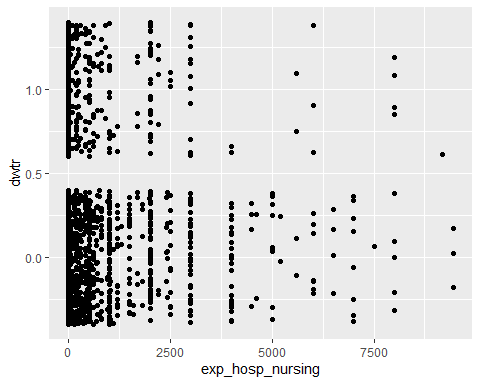


ggplot(dehn\_dat, aes(ehn, aehn)) + geom\_col() + y\_scale

## Warning: Removed 1 rows containing missing values (position\_stack).



#ggplot(ehn\_dat, aes(ehn, aehn, color = G)) + geom\_point(size = 2)  
  
dcom1 %>%   
 filter(exp\_hosp\_nursing < 10000, age > 64) %>%   
 ggplot(aes(exp\_hosp\_nursing, dwtr)) + geom\_jitter()



Whether Employed

demp <- c()  
femp <- c()  
memp <- c()  
emp <- c(1:6)  
  
for(i in c(1:6)) {  
 d6 <- filter(dcom1, Remploy == i, age > 64)  
 dM <- filter(dcom1, Remploy == i, age > 64, sex == 1)  
 dF <- filter(dcom1, Remploy == i, age > 64, sex == 2)  
 demp <- append(demp, sum(d6$dwtr) / length(d6$Remploy))  
 femp <- append(femp, sum(dF$dwtr) / length(dF$Remploy))  
 memp <- append(memp, sum(dM$dwtr) / length(dM$Remploy))  
}  
  
  
demp\_dat <- data.frame(emp, aemp = demp, G = "A")  
memp\_dat <- data.frame(emp, aemp = memp, G = "M")  
femp\_dat <- data.frame(emp, aemp = femp, G = "F")  
emp\_dat <- full\_join(memp\_dat, full\_join(femp\_dat, demp\_dat))

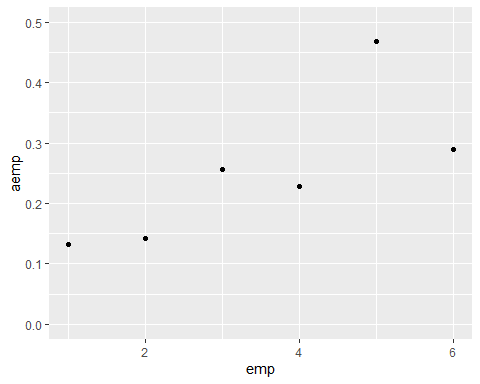
## Joining, by = c("emp", "aemp", "G")

## Warning: Column `G` joining factors with different levels, coercing to  
## character vector

## Joining, by = c("emp", "aemp", "G")

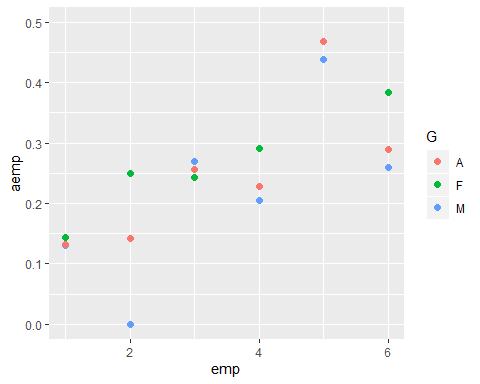
## Warning: Column `G` joining factor and character vector, coercing into  
## character vector

emp\_dat$G <- as.factor(emp\_dat$G)  
  
  
ggplot(demp\_dat, aes(emp, aemp)) + geom\_point() + y\_scale

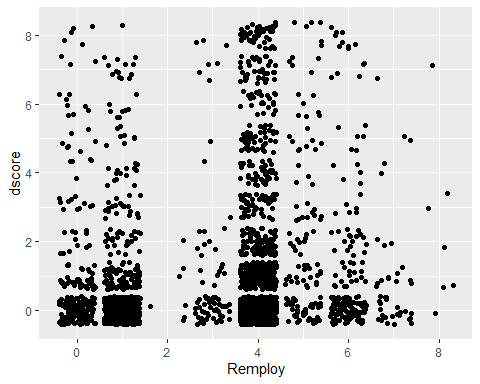


ggplot(emp\_dat, aes(emp, aemp, color = G)) + geom\_point(size = 2) + y\_scale

## Warning: Removed 1 rows containing missing values (geom\_point).



dcom1 %>%   
 filter(Remploy < 9, age > 64) %>%   
 ggplot(aes(Remploy, dscore)) + geom\_jitter()



Graph edits: - labels & titles