Project: Examining Dementia

Isabelle Sanford

1/14/2020

A. HISTORIES

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

B. PRESENT LIFE

i. SCORES

a. General Health
b. Life Satisfaction
c. Amount of Mental Activity
d. Rate of Interaction
e. Time Pressure

ii. TIME AMOUNTS

a. Hours of Personal Care
b. Hours of Leisure Time
c. Frequency of Attending Religious Services

C. POTENTIAL ISSUES

a. Age
b. Time Amounts - Binning

D. CONCLUSION

~

INTRODUCTION

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.

A. 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.

a. 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.

b. 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.)

c. 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).

d. 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.

B. PRESENT LIFE

Each of the below variables are about the individual's life as it presently is.

i. 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.

a. 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.)

b. 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%.

c. 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.

d. 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.

e. 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.

ii. 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).

a. 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.

b. 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.

c. 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.)

C. POTENTIAL ISSUES

a. 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

b. Binning - Leisure Hours

Here, the graph had to have 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.

D. CONCLUSION

Dementia has some correlation with degree of highest education, possibly as a proxy for economic status, but other historical factors examined do not seem to affect dementia rates. People with dementia also tend to rate their general health, life satisfaction and mental activity levels as relatively low. They have less interaction with others than average, spend less time on personal care, and more time in leisure activities. None of this information is especially new or groundbreaking, but it is still useful: in examining points of interest to look into further, and even more in underscoring the quality of life for people with dementia. This is why it is important to catch early, to find a cure for: because individuals with dementia have poorer health and are less happy with their lives.

Further research topics could include trying to remove age as a factor from other variables, further examining education levels alongside socioeconomic data, and a closer look at whether religion affects dementia rates at all.

Source: Panel Study of Income Dynamics at the University of Michigan. https://psidonline.isr.umich.edu/default.aspx To reproduce this exact dataset, the first column of the attached "Labels2.csv" contains the list of variables used.

Setup

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

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

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
d1 <- read_csv("Data2.csv") # read in the actual data
```

```
## Parsed with column specification:
## cols(
##   .default = col_double()
## )
```

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

```
labels <- read_csv("Labels2.csv") # read in the variable names
```

```
## Parsed with column specification:
## cols(
##   Number = col_character(),
##   Variable = col_character(),
##   New = col_character()
```

```
## )
```

Labeling and Cleaning

```
# rename variables using labels dataframe
labels <- filter(labels, is.na(New) == FALSE)
names(d1) <- labels$New

# remove unnecessary variables
d2 <- select(d1,-delete1, -delete2, -delete, -delete3, -delete4, -delete5)

# remove families with neither head nor spouse older than 65
d3 <- filter(d2, R65plus == 1 | S65plus == 1)
```

Combining Head & Spouse data

```
# create a dataset of only data about the head of house, removing 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))

# create a spouse-only dataset and rename spouse-specific variables to neutral
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)

# rename head-spectific variables to neutral
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)

#combine the two sets so each line is about either a head or a spouse
dcom <- full_join(dR, dS)
```

```
## Joining, by = c("FID68", "PID68", "FID17", "PID17", "state", "age", "sex", "retiredwhen", "hrs_perso
```

Education & Dementia scores

```
# combining various education variables into one "educ" variable
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))))))
# key:
# 7 = PhD
# 6 = Master's / law or medical degree
# 5 = Bachelors, 4 = Associates
# 3 = Some college but unfinished
# 2 = high school grad or GED
# 1 = did not finish high school


# combining all dementia-related variables into a single score between 0 and 8
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)

# remove education and dementia variables that are now redundant
dcom1 <- dcom %>%
  select(-c(c_interest, c_story_repeats, c_learning, c_datemem, c_money, c_apptmem, c_memory, c_decisio
  select(-c(wtr_hsgrad, ged_grade, hs_grade, wtr_uni, uni_grade, wtr_unigrad, highest_degree)) %>%
  mutate(dscore = dscore, educ = educ)

# create a 'dwtr' variable which is a binary dementia-or-not variable (defined as 2+ on the dscore)
dwtr <- ifelse(dscore > 1, 1, 0)
dcom1 <- mutate(dcom1, dwtr = dwtr)
```

Graph Axis Prep

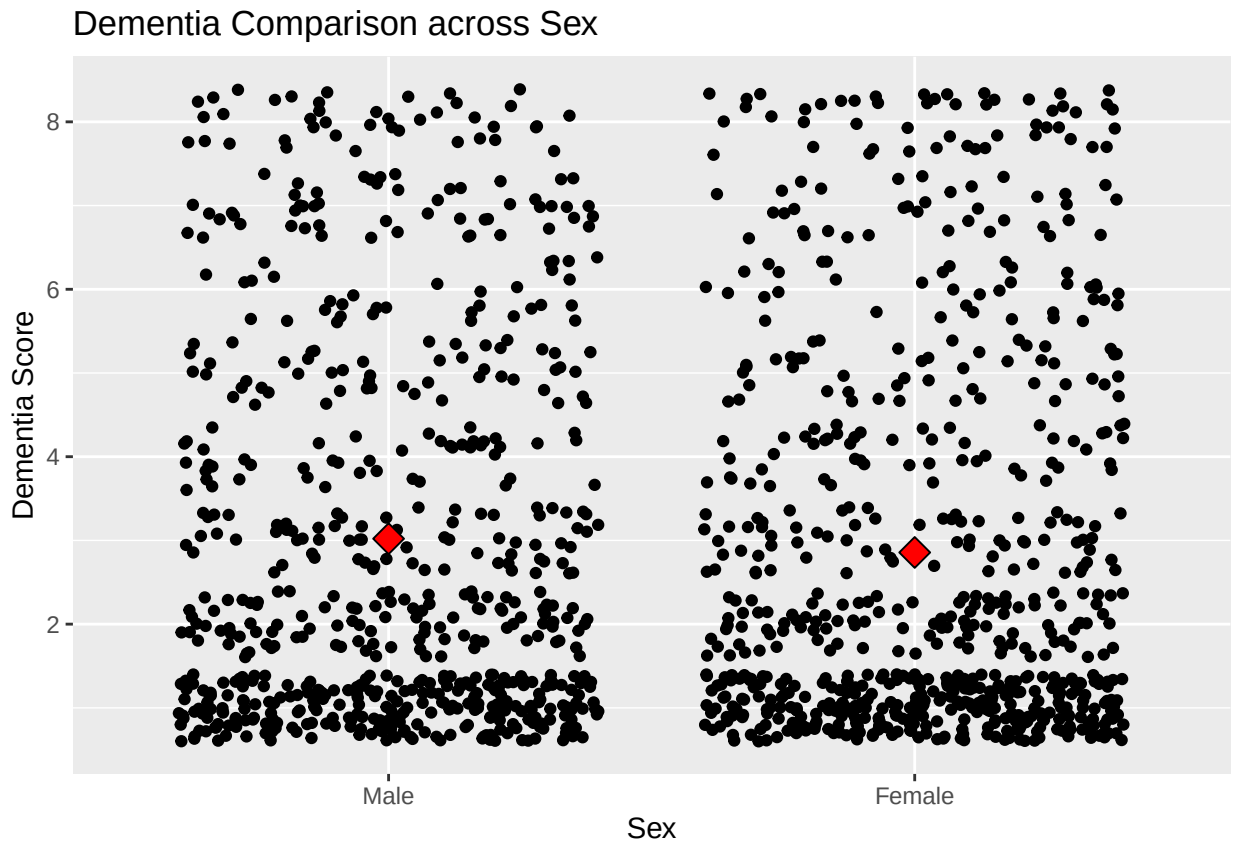
```
# y scale for use in graphing later, to make all relevant graphs have the same scale & start at 0
y_scale <- scale_y_continuous(limits = c(0,.5))
```

~

A. HISTORY VARIABLES

a. D vs Sex

```
dcom1$sex <- as.factor(dcom1$sex)

# graph male and female scores beside each other
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") + # show mean value
    scale_x_discrete(labels = c("Male", "Female")) + # in the data 1 is male and 2 is female
    labs(x = "Sex", y = "Dementia Score",
       title = "Dementia Comparison across Sex")
```




```
# effectively shows the percent of males and females respectively without dementia
# (i.e. the dimension(length) of the filtered set with only men without dementia
# divided by the dimension(length) of the set of all men)
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))
```

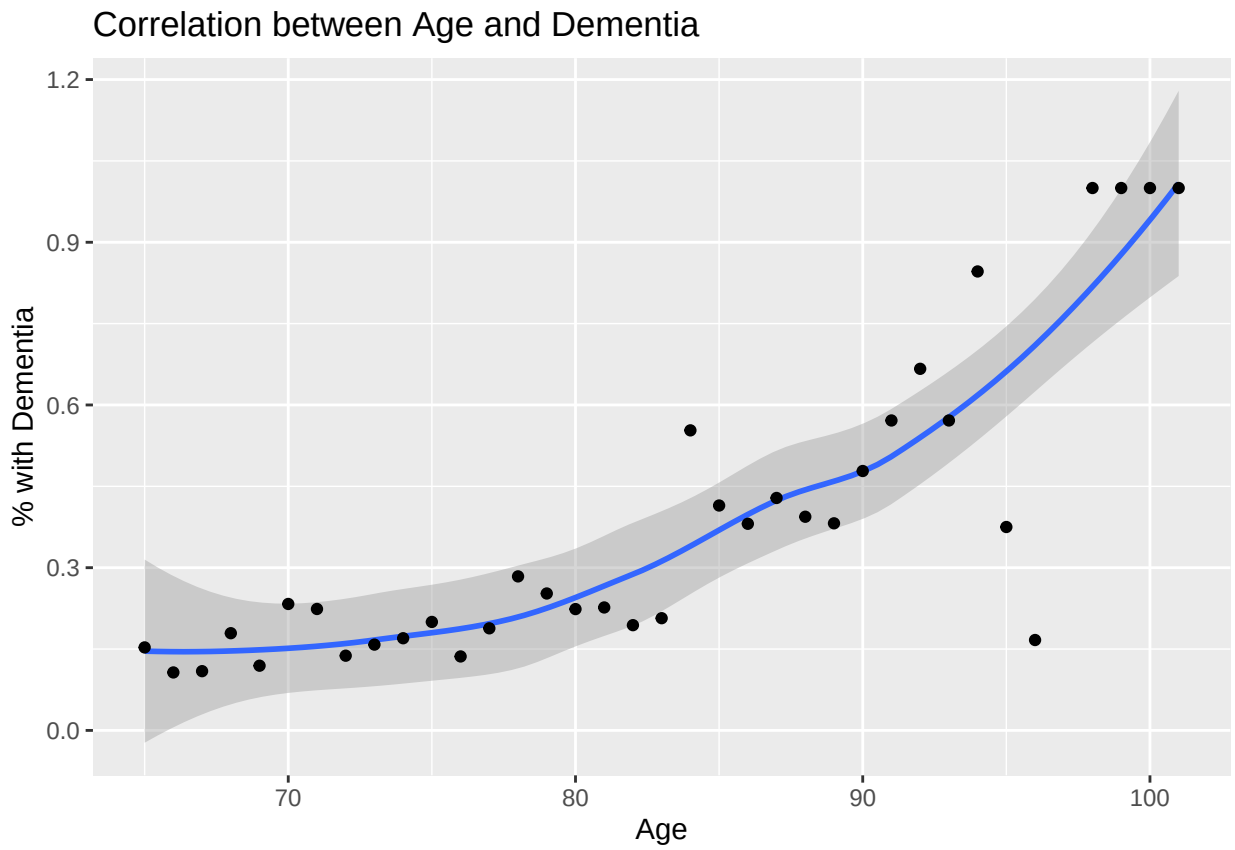
```
## [1] 0.8132586 1.0000000
```

The red diamonds indicate the mean score for each sex, and the two numbers (.80 and .81) indicate percent of individuals who do not have a score of 2 or above.

b. Dementia vs Age

```
# Most sections below this are very similar in structure.
# This section is commented in full detail
# and the rest are only commented where there's some deviation to note

# use the summarize function to create a dataset with the percent of people with dementia at each age
# (i.e. mean(dwtr), as dwtr is 1 when the individual has dementia and 0 otherwise)
a <- dcom1 %>%
  filter(age > 64) %>%
  group_by(age) %>%
  summarize(per = mean(dwtr))
# note: summary data isn't included here for brevity as the graphs are generally sufficient
# but running the appropriate letter (a - k) will show it if interested

# plot the table just created on a graph, with points and a geom_smooth() where applicable
ggplot(a, aes(age, per)) +
  geom_smooth() +
  geom_point() +
  labs(x = "Age", y = "% with Dementia",
       title = "Correlation between Age and Dementia")
```

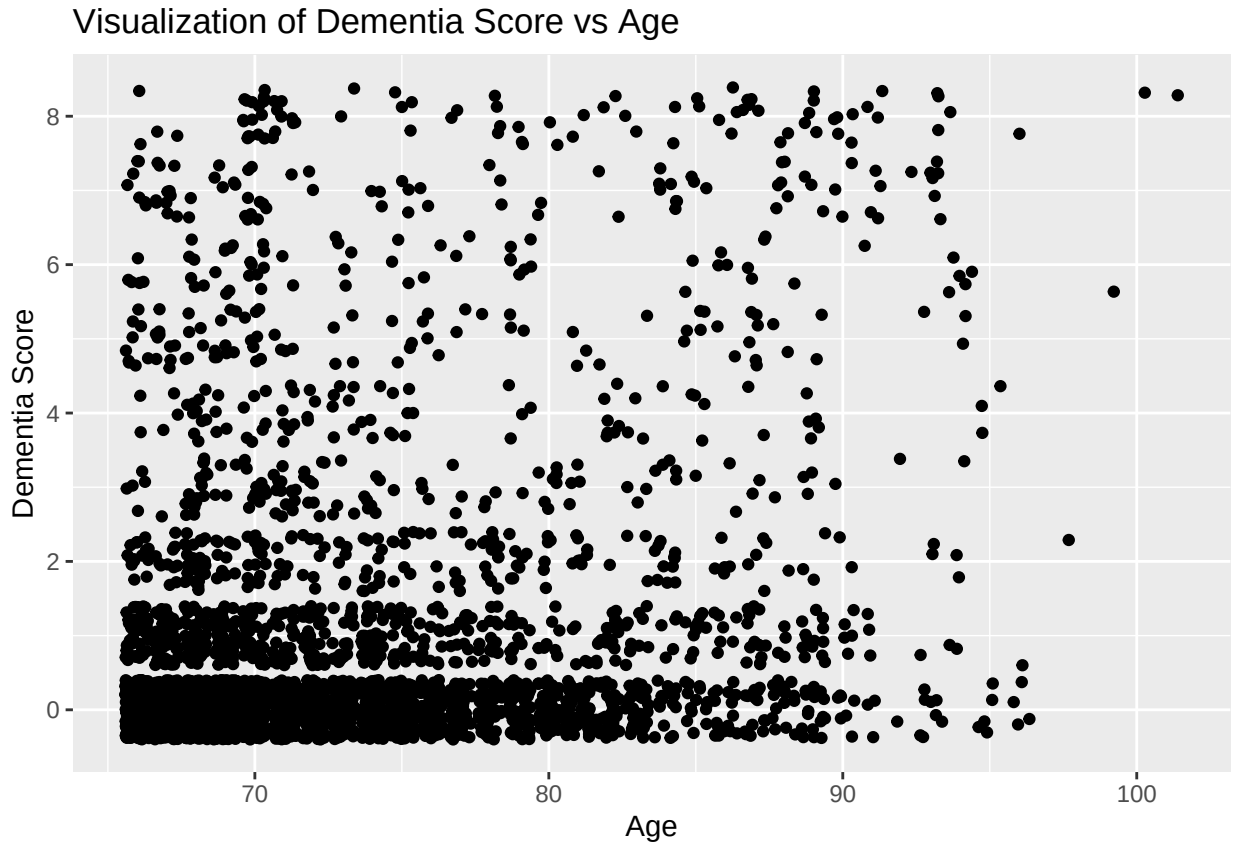
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
# calculate correlation coefficient between age and dementia rate
cor(a$age, a$per)
```

```
## [1] 0.8245455
```

```
# using the original dataset, plot age vs the dementia score
# to get a better notion of absolute numbers and where variance might be due to low sample size
dcom1 %>%
  filter(age > 65) %>%
  ggplot(aes(age, dscore)) +
    geom_jitter() +
    labs(x = "Age", y = "Dementia Score",
       title = "Visualization of Dementia Score vs Age")
```

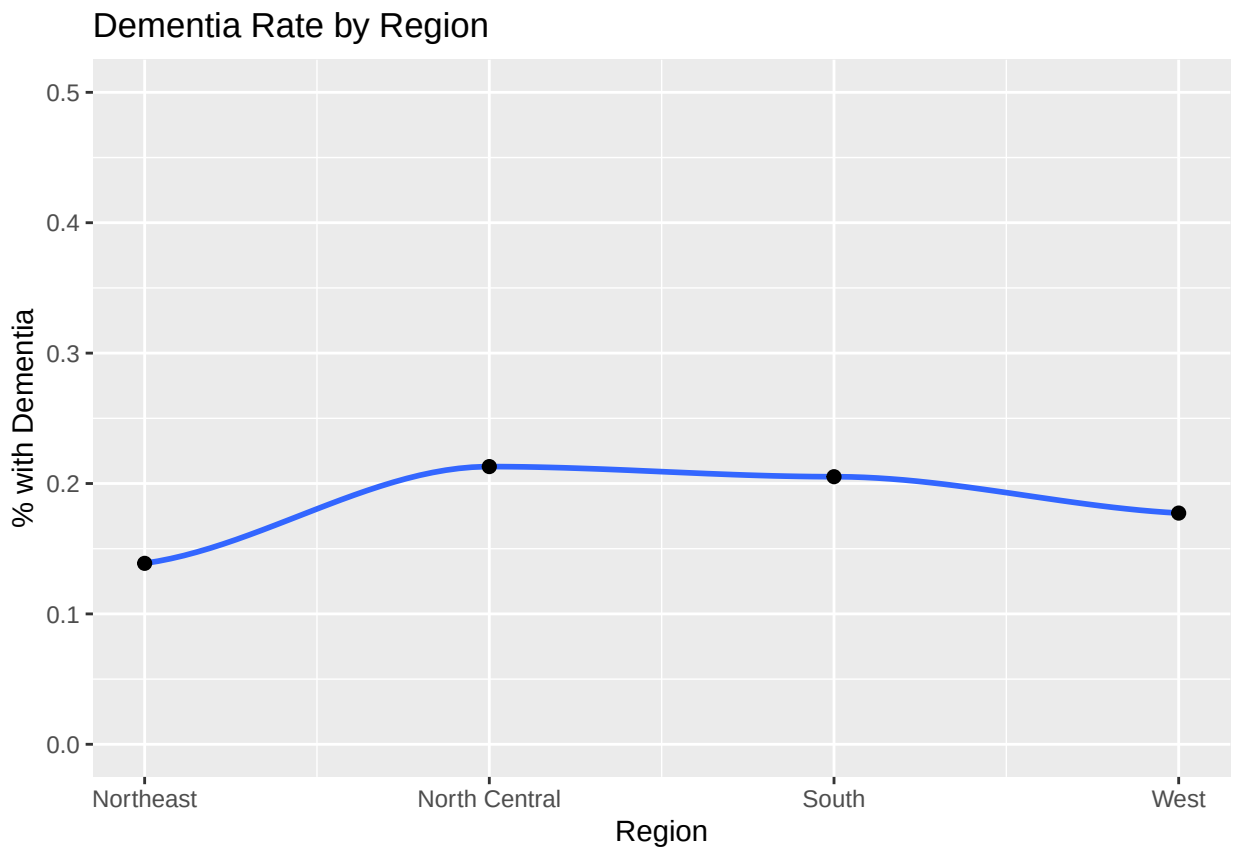


c. D vs Region

```
b <- dcom1 %>%
# many variables have a NA / dead option, which is the reason for almost all non-age filters
  filter(age > 64, region < 5) %>%
  group_by(region) %>%
  summarize(n(), per = mean(dwtr))

ggplot(b, aes(region, per)) +
  geom_smooth(na.rm = T) + # otherwise lots of warnings
  geom_point(size = 2) +
  y_scale +
  scale_x_continuous(
    labels = c("Northeast",  "North Central", "South", "West")) + # 1,2,3, and 4 in the data
  labs(x = "Region", y = "% with Dementia",
       title = "Dementia Rate by Region")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

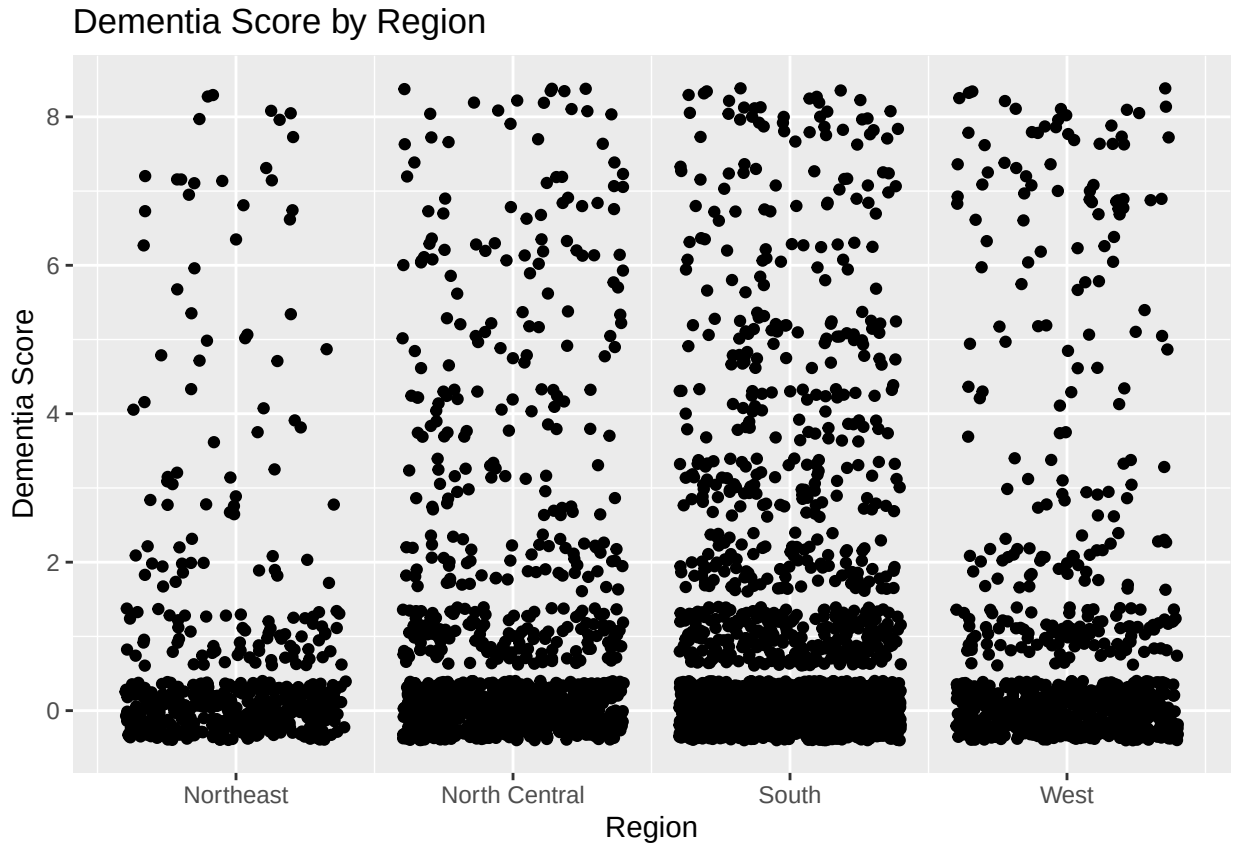
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 0.985

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 2.015

## 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. 4.0602

## 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 0.985

## 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
## 2.015

## 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. 4.0602
```




```
cor(b$region, b$per)
```

```
## [1] 0.4149174
```

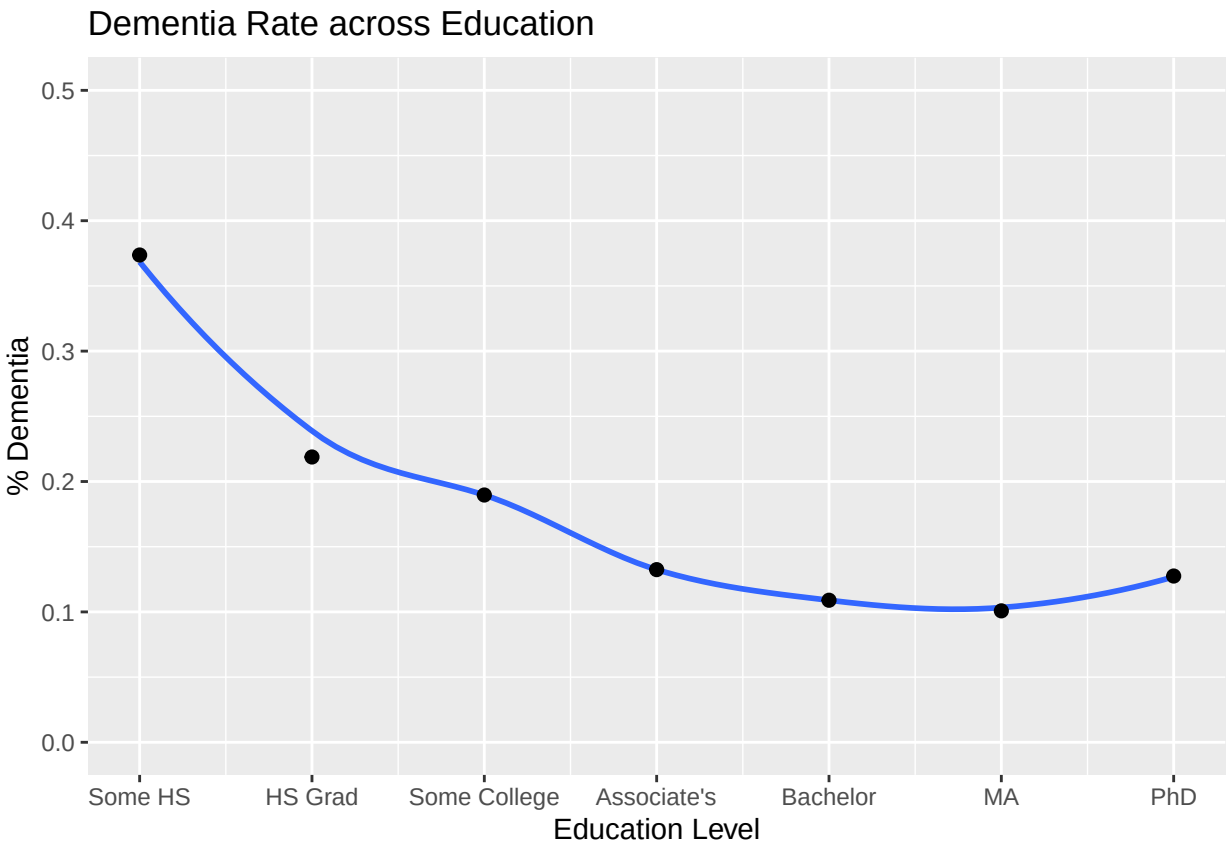
```
dcom1 %>%
  filter(region < 5, age > 64) %>%
  ggplot(aes(region, dscore)) +
    geom_jitter() +
    scale_x_continuous(
      labels = c("Northeast",  "North Central", "South", "West"),
      breaks = c(1:4)) +
  # the breaks function is there because for some reason without it R thinks there aren't four labels
  # (even though there absolutely are)
    labs(x = "Region", y = "Dementia Score",
       title = "Dementia Score by Region")
```



d. D vs Education

```
c <- dcom1 %>%
  filter(age > 64) %>%
  group_by(educ) %>%
  summarize(n(), per = mean(dwtr))

ggplot(c, aes(educ, per)) +
  geom_smooth(se = F) +
  geom_point(size = 2) +
  y_scale  +
  scale_x_continuous(
    labels = c("Some HS", "HS Grad", "Some College", "Associate's", "Bachelor", "MA", "PhD"),
    breaks = c(1:7)) +
  labs(x = "Education Level", y = "% Dementia",
       title = "Dementia Rate across Education")
```

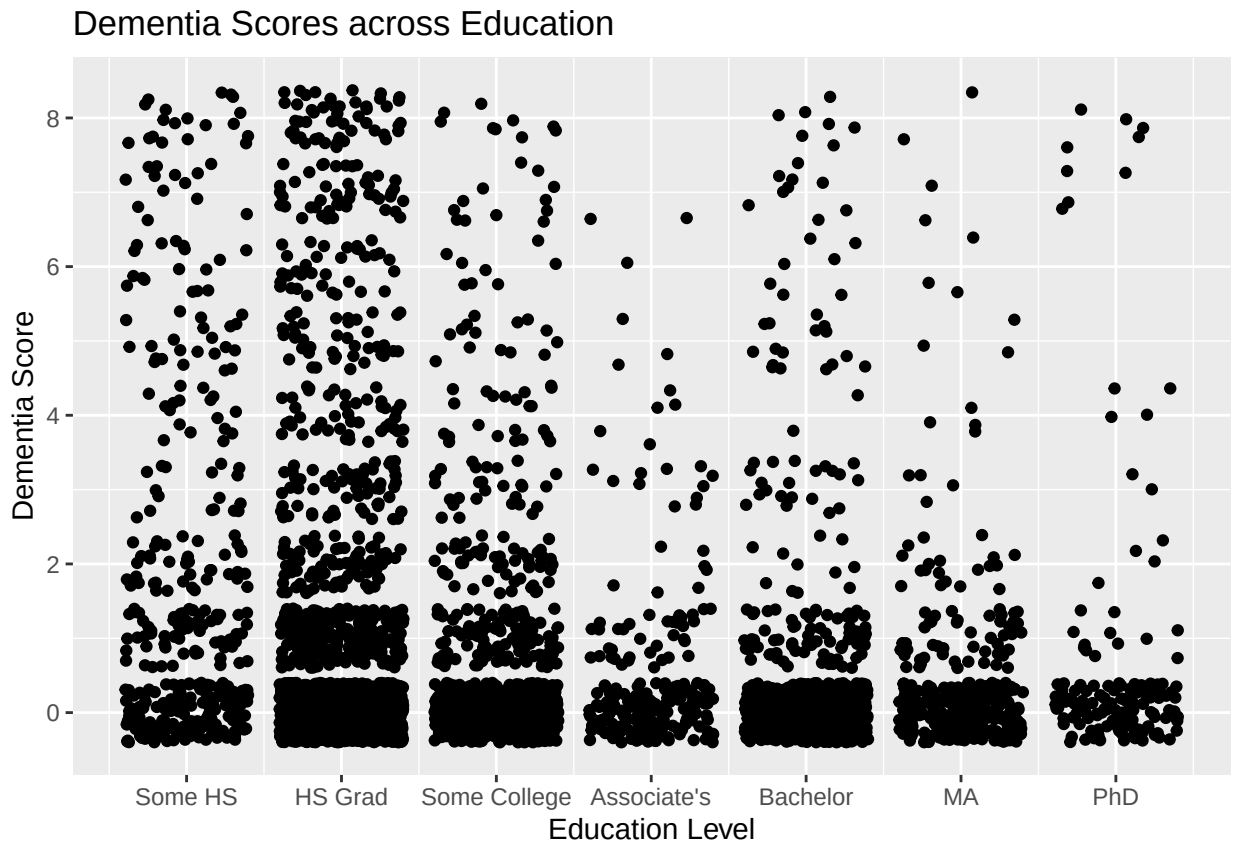
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
cor(c$educ, c$per)
```

```
## [1] -0.8469022
```

```
dcom1 %>%
  filter(educ < 8, age > 64) %>%
  ggplot(aes(educ, dscore)) +
    geom_jitter() +
    scale_x_continuous(
      labels = c("Some HS", "HS Grad", "Some College", "Associate's", "Bachelor", "MA", "PhD"),
      breaks = c(1:7)) +
    labs(x = "Education Level", y = "Dementia Score",
       title = "Dementia Scores across Education")
```

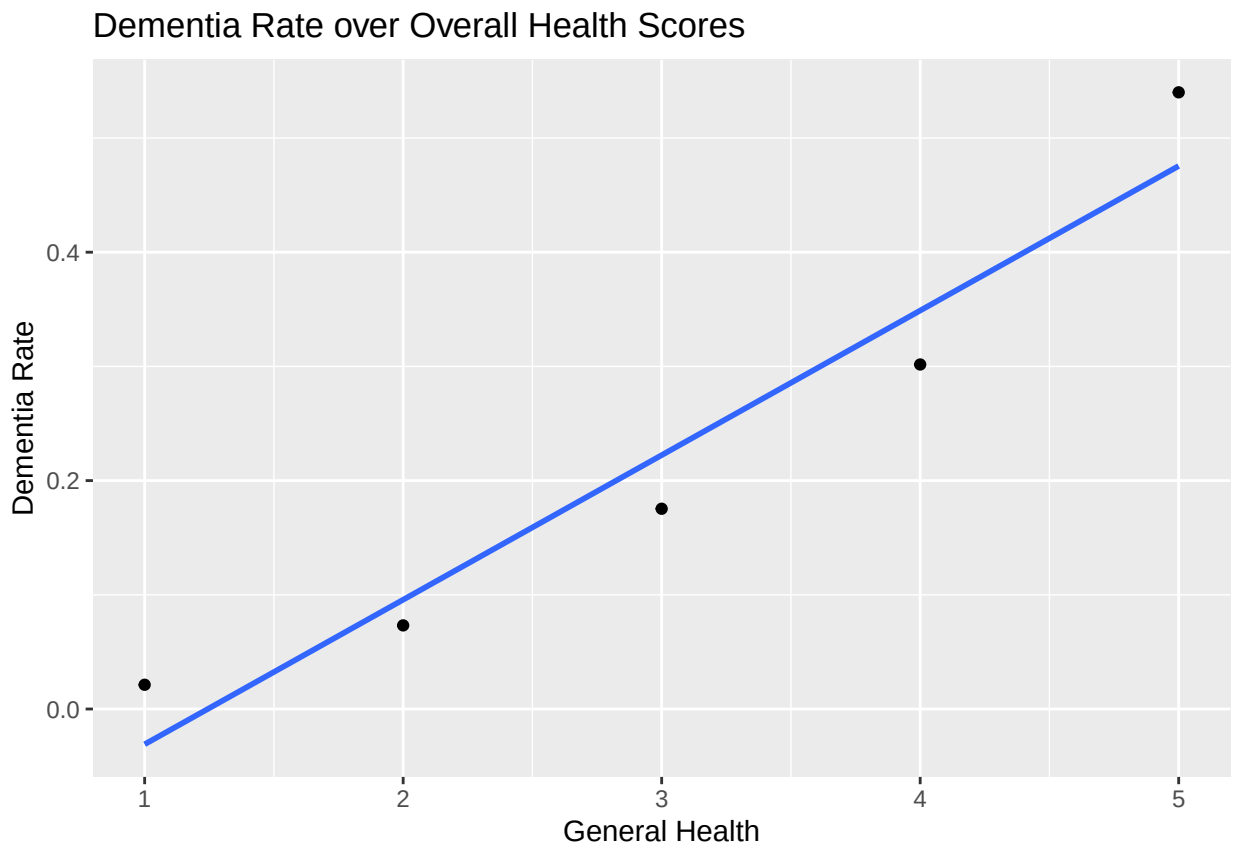


A. PRESENT LIFE

i. SCORES

a. General Health

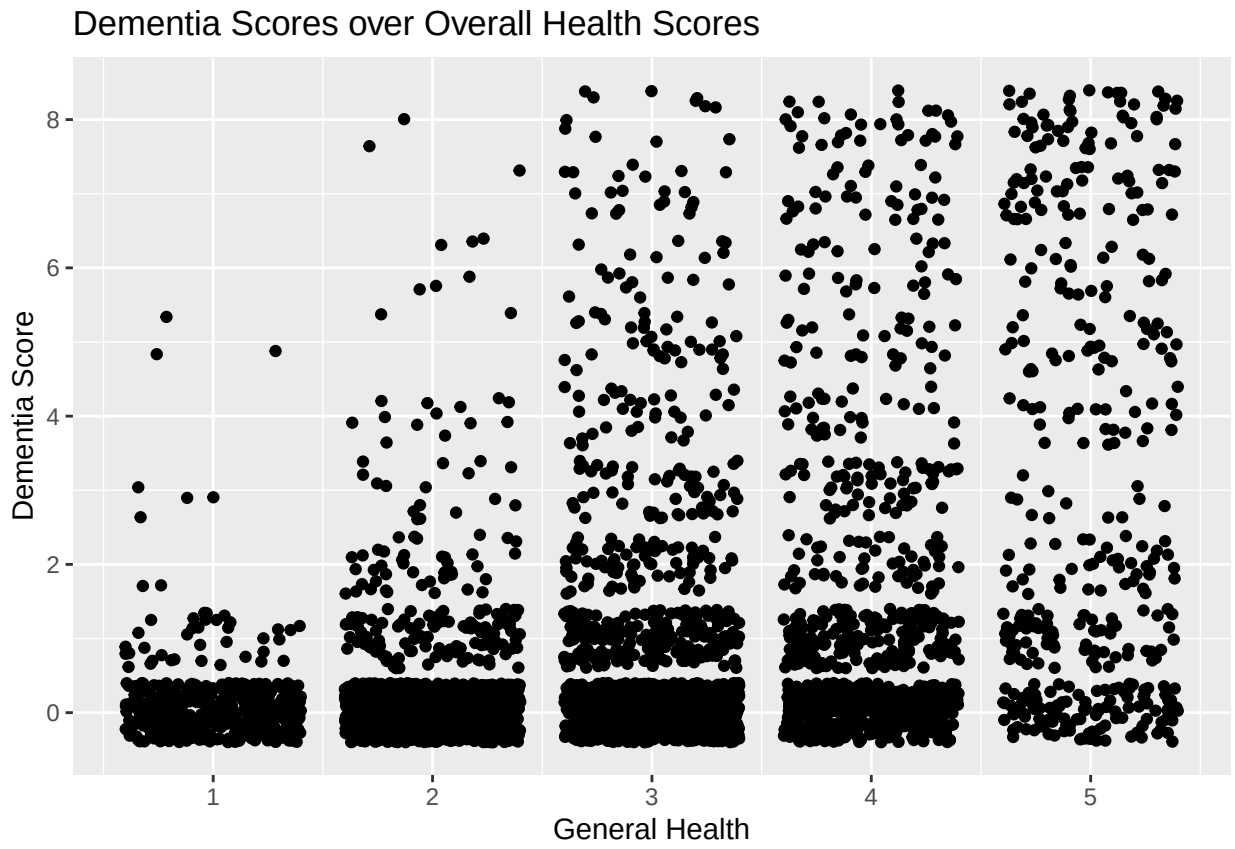
```
d <- dcom1 %>%
  filter(age > 64, health_general < 6) %>%
  group_by(health_general) %>%
  summarize(n(), per = mean(dwtr))

ggplot(d, aes(health_general, per)) +
  geom_point() +
  geom_smooth(method = "lm", se = F) +
  labs(x = "General Health", y = "Dementia Rate",
       title = "Dementia Rate over Overall Health Scores")
```




```
cor(d$health_general, d$per)
```

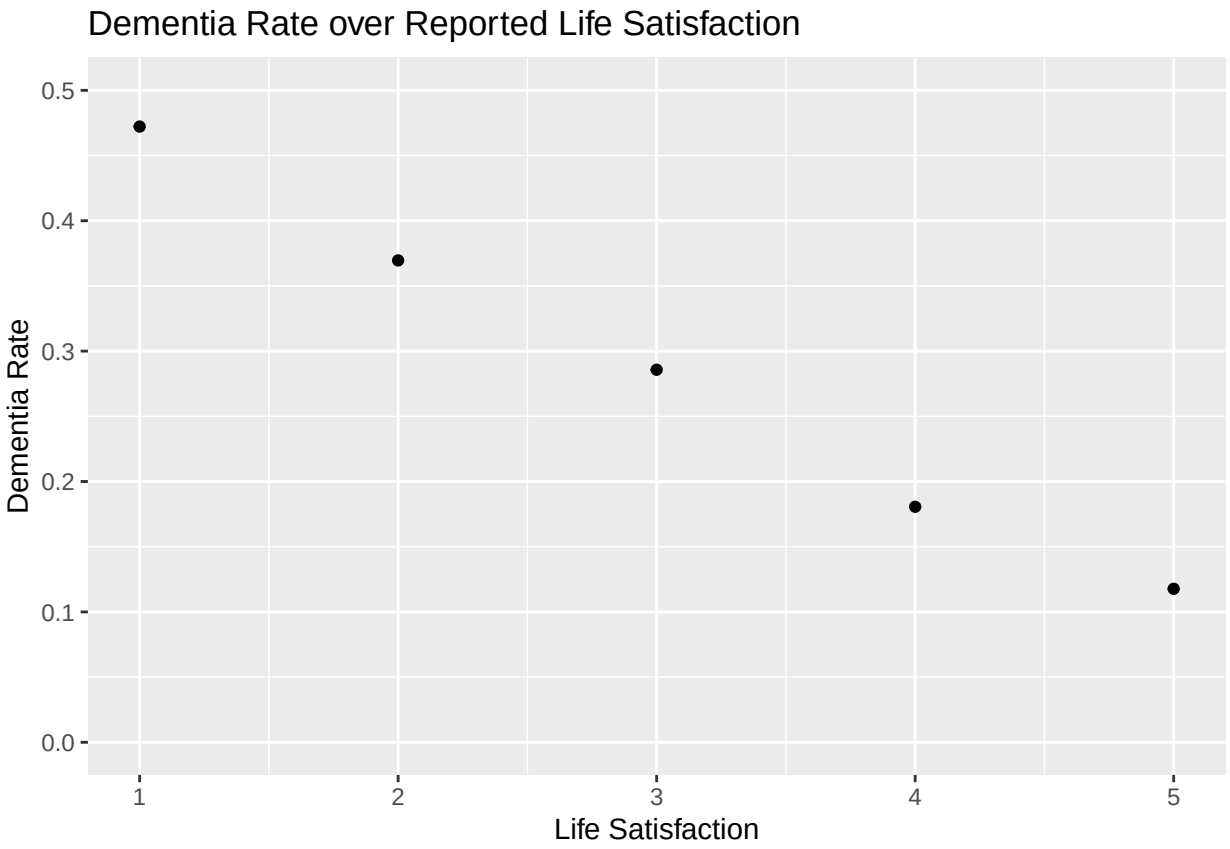
```
## [1] 0.9650608
```

```
dcom1 %>%
  filter(health_general < 6, age > 64) %>%
  ggplot(aes(health_general, dscore)) +
    geom_jitter() +
    labs(x = "General Health", y = "Dementia Score",
       title = "Dementia Scores over Overall Health Scores")
```



b. D vs Life Satisfaction

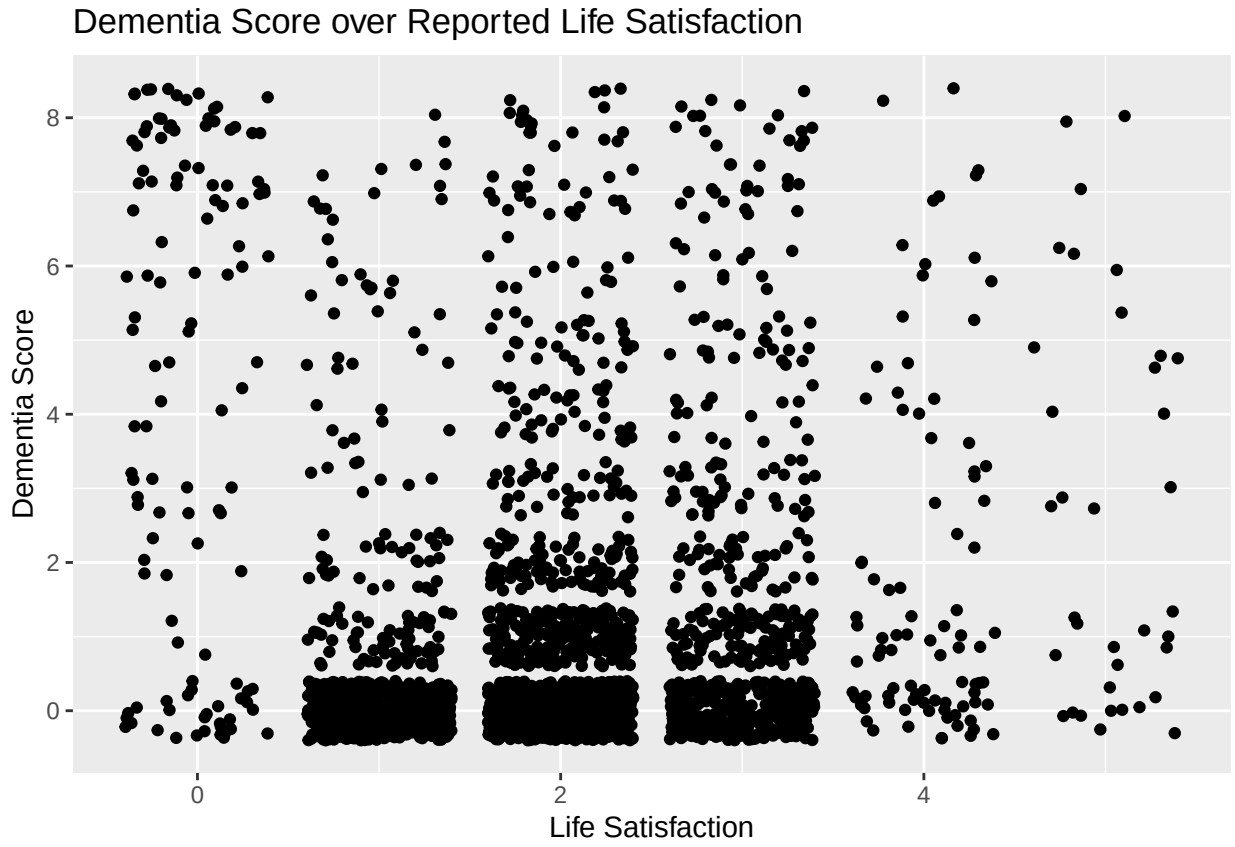
```
e <- dcom1 %>%
  filter(age > 64, life_satisf < 6, life_satisf > 0) %>%
# this reverses the scale for life satisfaction, as it originally had 1 as best and 5 as worst
  group_by(lifesatisf = -(life_satisf - 3) + 3) %>%
  summarize(n(), per = mean(dwtr))

ggplot(e, aes(lifesatisf, per)) +
  geom_point() +
  y_scale +
  labs(x = "Life Satisfaction", y = "Dementia Rate",
       title = "Dementia Rate over Reported Life Satisfaction")
```




```
cor(e$lifesatisf, e$per)
```

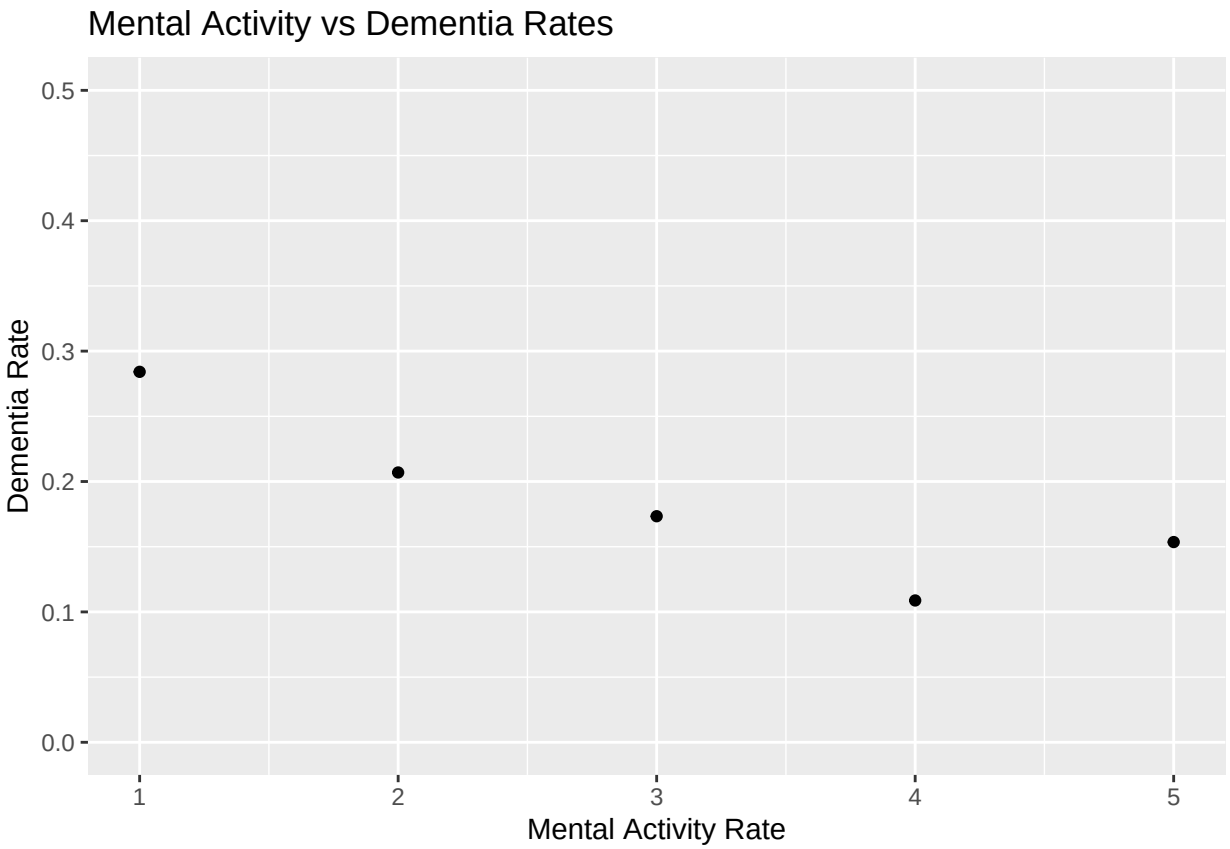
```
## [1] -0.9972192
```

```
dcom1 %>%
  filter(life_satisf < 6, age > 64) %>%
  ggplot(aes(life_satisf, dscore)) +
    geom_jitter() +
    labs(x = "Life Satisfaction", y = "Dementia Score",
       title = "Dementia Score over Reported Life Satisfaction")
```



c. D vs Mental Activity

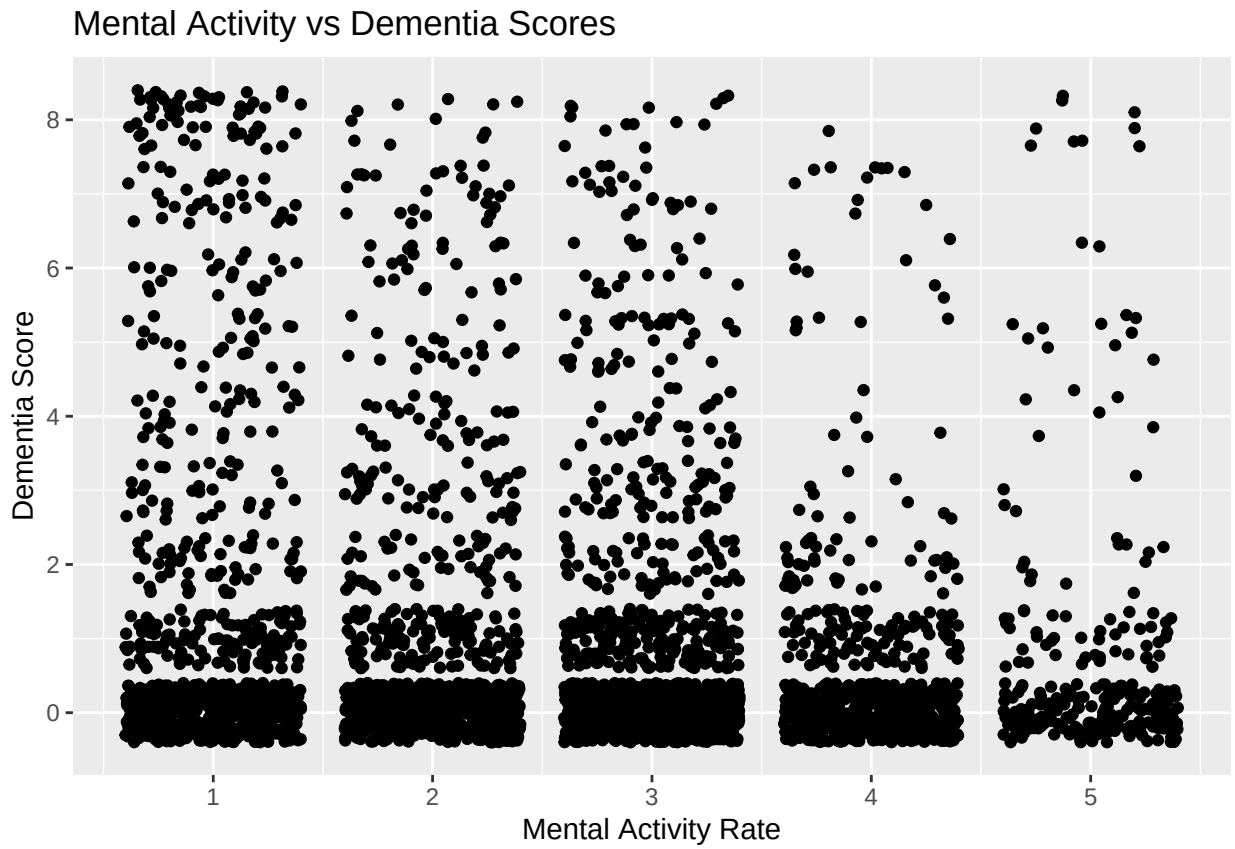
```
f <- dcom1 %>%
  filter(age > 64, mental_activity < 6, mental_activity > 0) %>%
  group_by(mental_activity) %>%
  summarize(n(), per = mean(dwtr))

ggplot(f, aes(mental_activity, per)) +
  geom_point() +
  y_scale +
  labs(x = "Mental Activity Rate", y = "Dementia Rate",
       title = "Mental Activity vs Dementia Rates")
```




```
cor(f$mental_activity, f$per)
```

```
## [1] -0.8652993
```

```
dcom1 %>%
  filter(mental_activity < 6, age > 64) %>%
  ggplot(aes(mental_activity, dscore)) +
    geom_jitter() +
    labs(x = "Mental Activity Rate", y = "Dementia Score",
       title = "Mental Activity vs Dementia Scores")
```



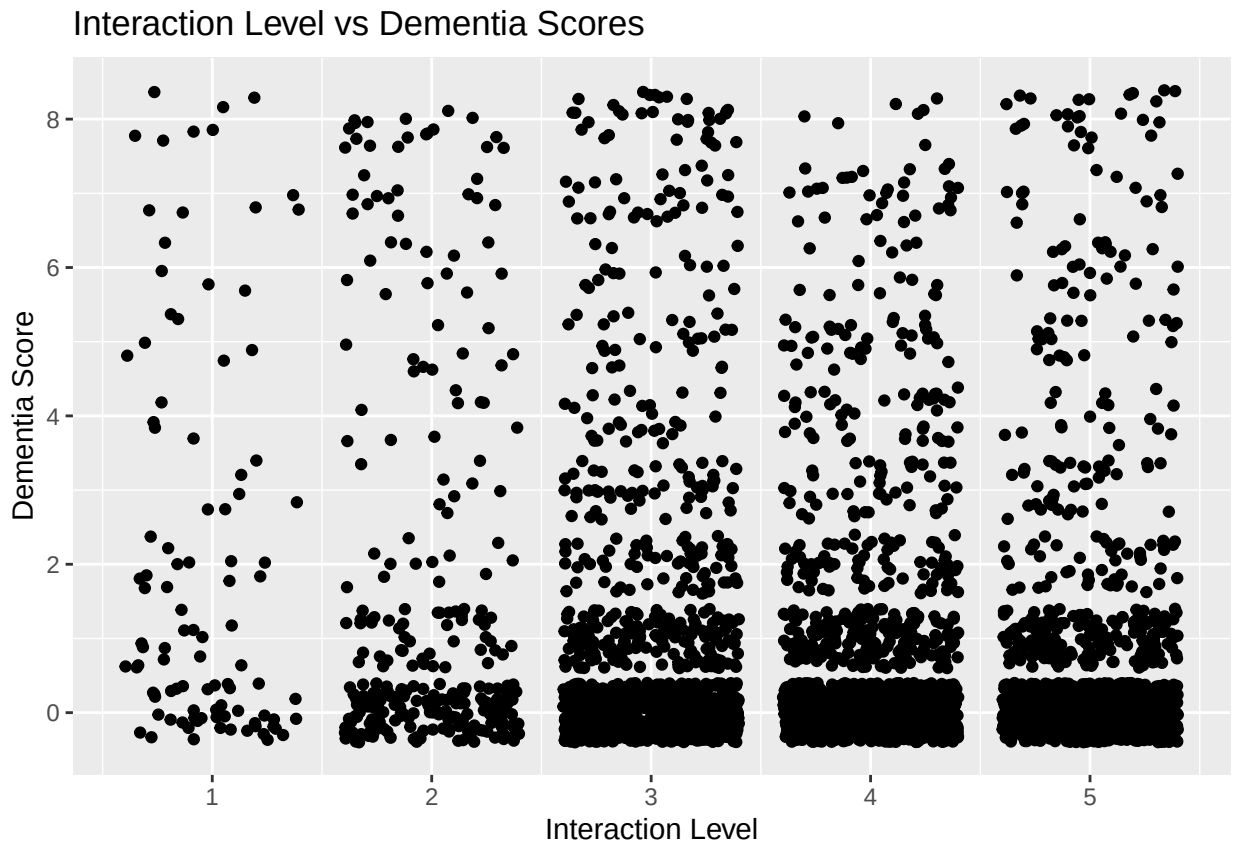
d. D vs Interaction

```
g <- dcom1 %>%
  filter(age > 64, interaction < 6) %>%
  group_by(interaction) %>%
  summarize(n(), per = mean(dwtr))

ggplot(g, aes(interaction, per)) +
  geom_point()+ y_scale +
  labs(x = "Interaction Level", y = "Dementia Rate",
       title = "Interaction Level vs Dementia Rates")
```


```
cor(g$interaction, g$per)
```

```
## [1] -0.9198645
```

```
dcom1 %>%
  filter(interaction < 6, age > 64) %>%
  ggplot(aes(interaction, dscore)) + geom_jitter()+
  labs(x = "Interaction Level", y = "Dementia Score",
       title = "Interaction Level vs Dementia Scores")
```



(bcd) Combination of previous 3

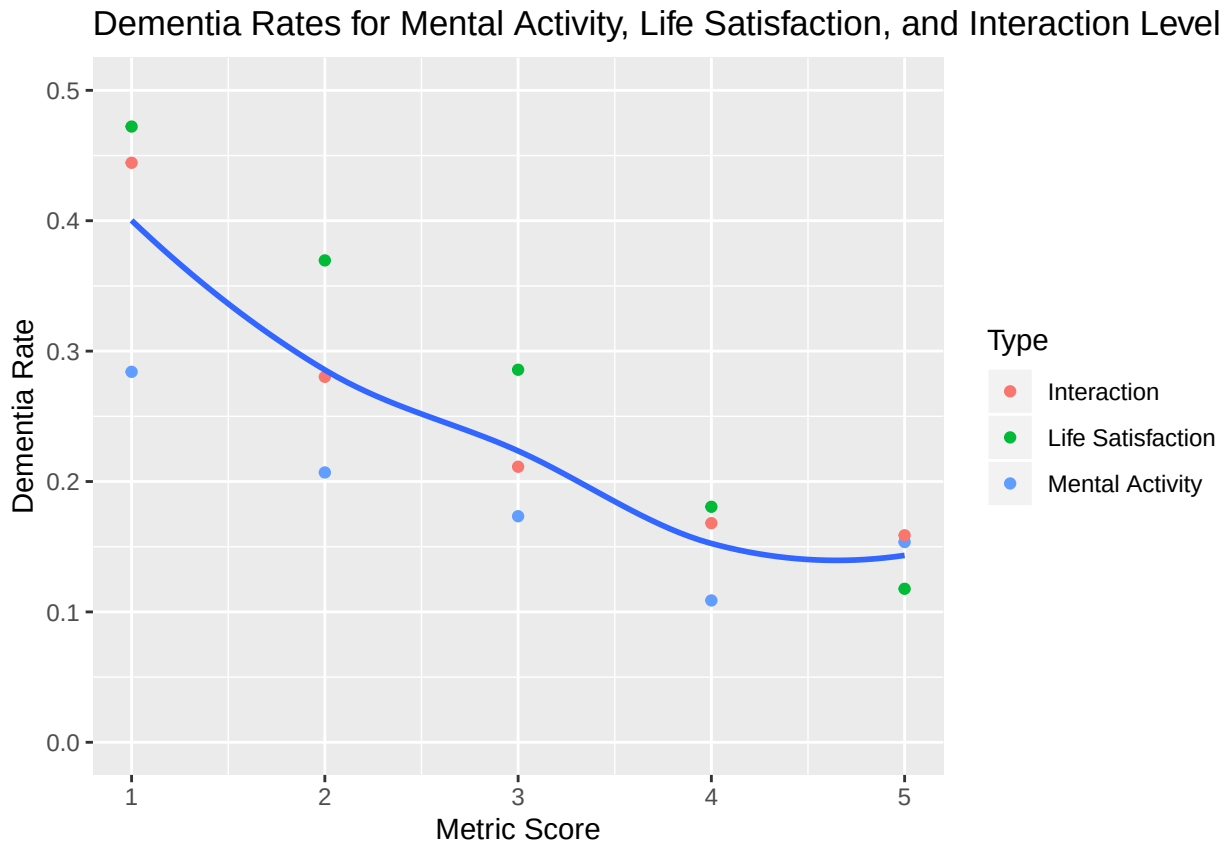
```
#combine e, f, and g into a single dataset, mutating by score (vec) and type for ease of graphing
dat3 <- e %>%
  mutate(vec = lifesatisf, Type = "Life Satisfaction") %>%
  full_join(mutate(f, vec = mental_activity, Type = "Mental Activity")) %>%
  full_join(mutate(g, vec = interaction, Type = "Interaction")) %>%
  select(per, vec, Type) # removes unnecessary columns
```

```
## Joining, by = c("n()", "per", "vec", "Type")
## Joining, by = c("n()", "per", "vec", "Type")
```

```
dat3$type <- as.factor(dat3$Type)
```

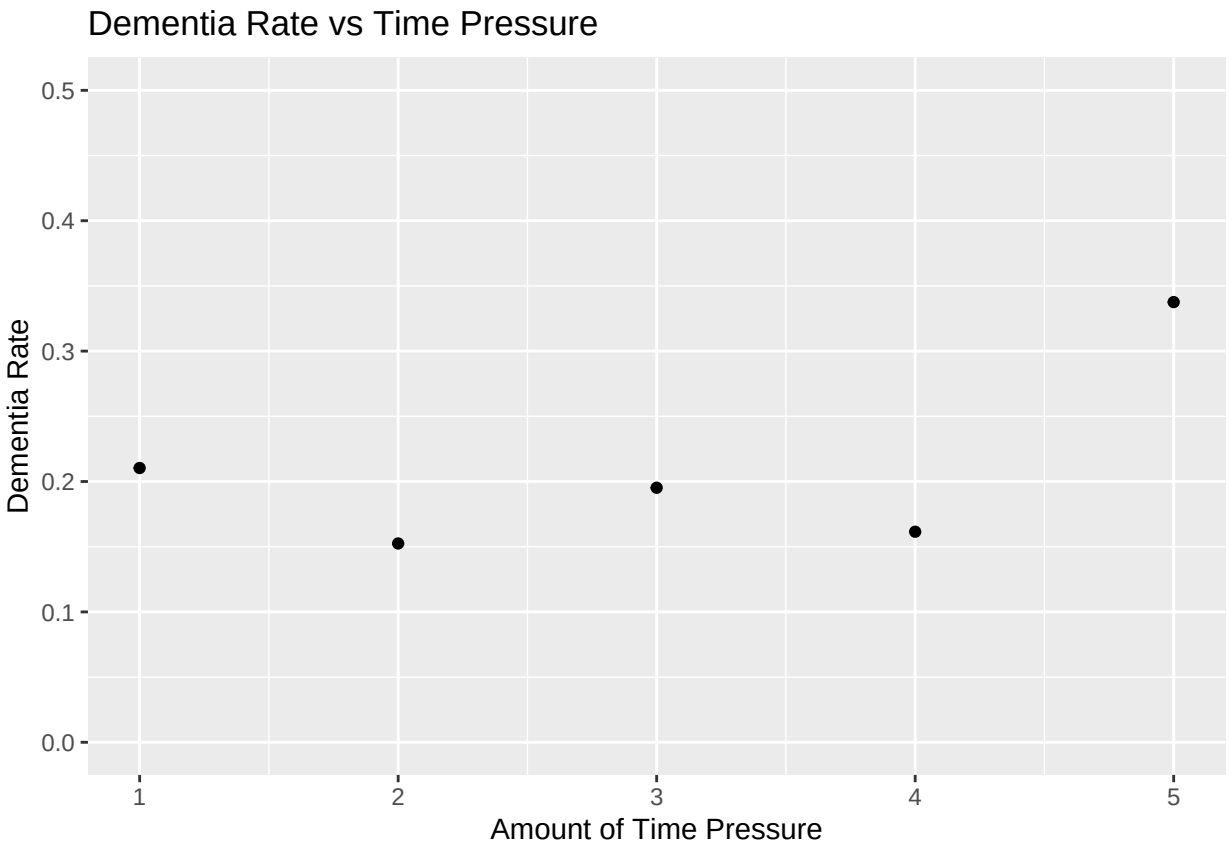
```
# plot all three on the same graph, with color
ggplot(dat3, aes(vec, per)) +
  geom_point(aes(color = Type)) +
  geom_smooth(se = F) +
  y_scale +
  labs(x = "Metric Score", y = "Dementia Rate",
       title = "Dementia Rates for Mental Activity, Life Satisfaction, and Interaction Level")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



e. D vs Time Pressure

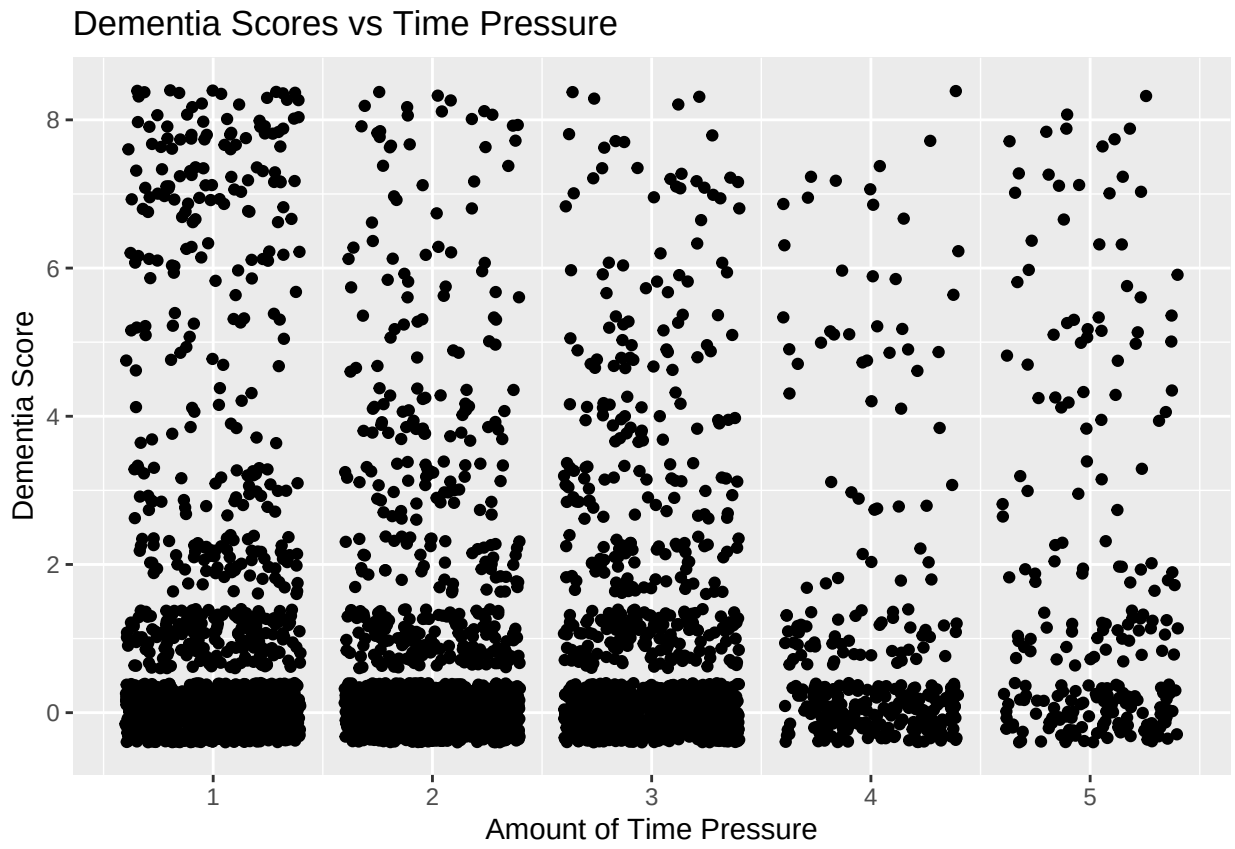
```
h <- dcom1 %>%
  filter(age > 64, timepressure < 6) %>%
  group_by(timepressure) %>%
  summarize(n(), per = mean(dwtr))

ggplot(h, aes(timepressure, per)) +
  geom_point() + y_scale +
  labs(x = "Amount of Time Pressure", y = "Dementia Rate",
       title = "Dementia Rate vs Time Pressure")
```




```
cor(h$timepressure, h$per)
```

```
## [1] 0.5598491
```

```
dcom1 %>%
  filter(timepressure < 6, age > 64) %>%
  ggplot(aes(timepressure, dscore)) +
    geom_jitter() +
    labs(x = "Amount of Time Pressure", y = "Dementia Score",
       title = "Dementia Scores vs Time Pressure")
```

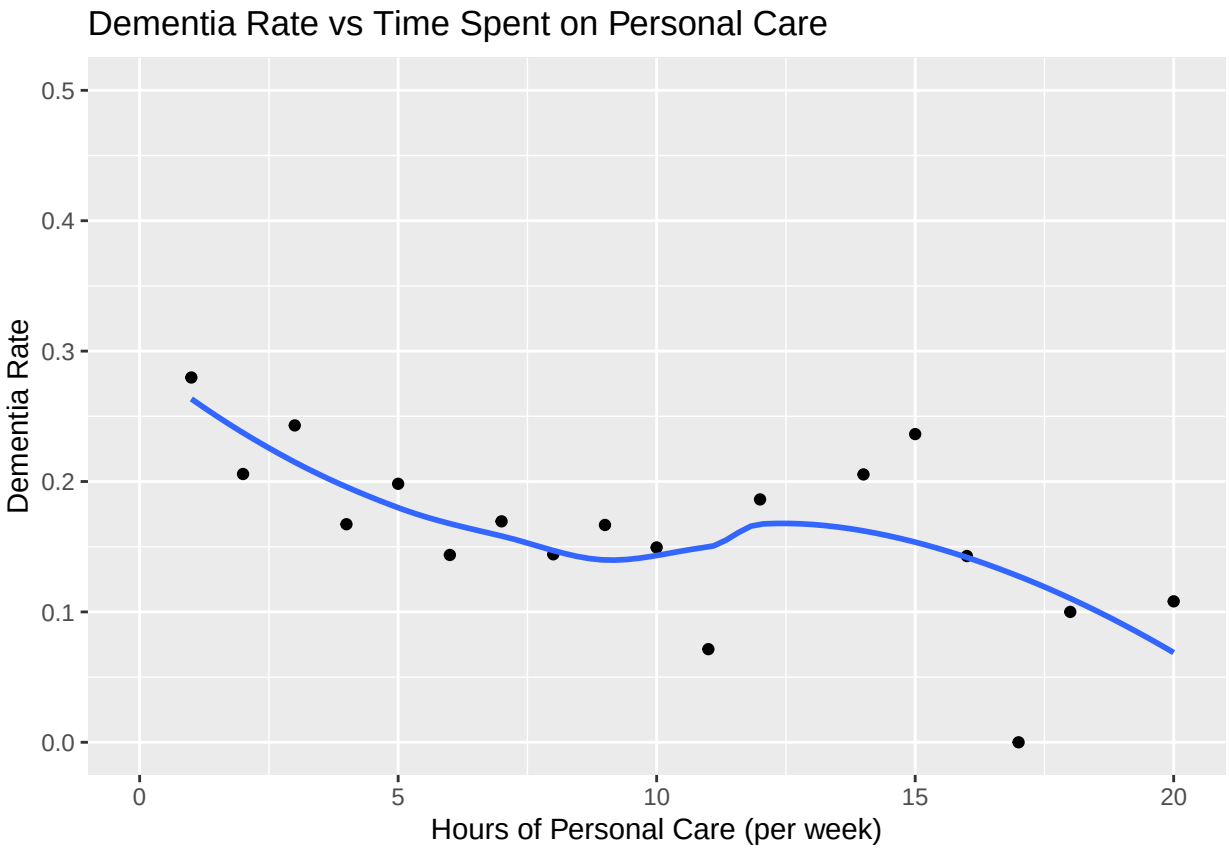


B. TIME AMOUNTS

a. D vs Personal Hours

```
 # there's somewhat more variance in section B because there are more than 5 possibilities
i <- dcom1 %>%
  filter(age > 64, hrs_personal < 21) %>%
  group_by(hrs_personal) %>%
  summarize(n(), per = mean(dwtr))

ggplot(i, aes(hrs_personal, per)) +
  geom_point() + y_scale + geom_smooth(se = F) +
  labs(x = "Hours of Personal Care (per week)", y = "Dementia Rate",
       title = "Dementia Rate vs Time Spent on Personal Care")
```

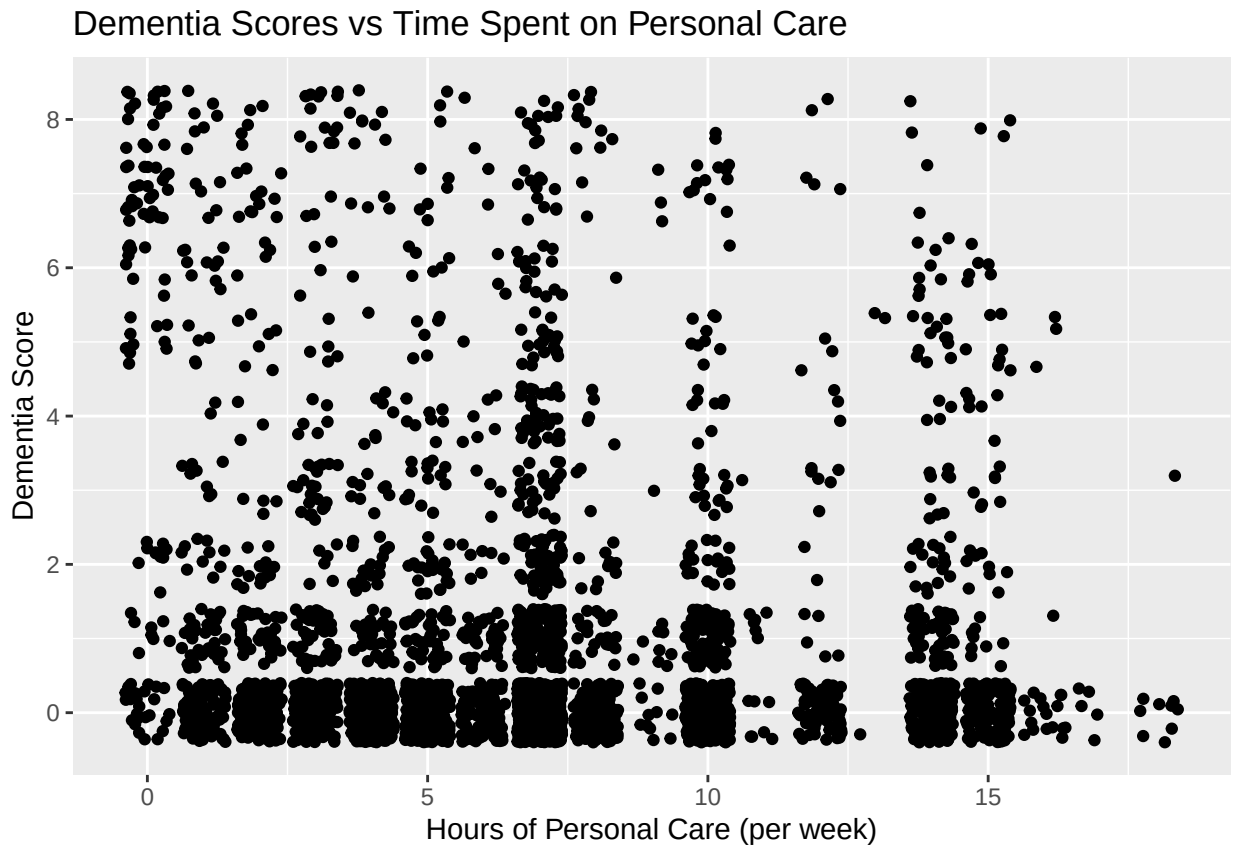
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
## Warning: Removed 2 rows containing missing values (geom_point).
```




```
cor(i$hrs_personal, i$per)
```

```
## [1] -0.3682538
```

```
dcom1 %>%
  filter(hrs_personal < 20, age > 64) %>%
  ggplot(aes(hrs_personal, dscore)) +
    geom_jitter() +
    labs(x = "Hours of Personal Care (per week)", y = "Dementia Score",
       title = "Dementia Scores vs Time Spent on Personal Care")
```



b. D vs Leisure Hours

```
# here, the technique used up to this point fails, as 63 bins makes for too much variance
# instead, bins are done by 5 to show a more useful picture

# same process as above here serves only to find the correlation coefficient (low due to variance)
j <- dcom1 %>%
  filter(age > 64, hrs_leisure < 64) %>%
  group_by(hrs_leisure) %>%
  summarize(n(), per = mean(dwtr))

cor(j$hrs_leisure, j$per)
```

```
## [1] -0.03188192
```

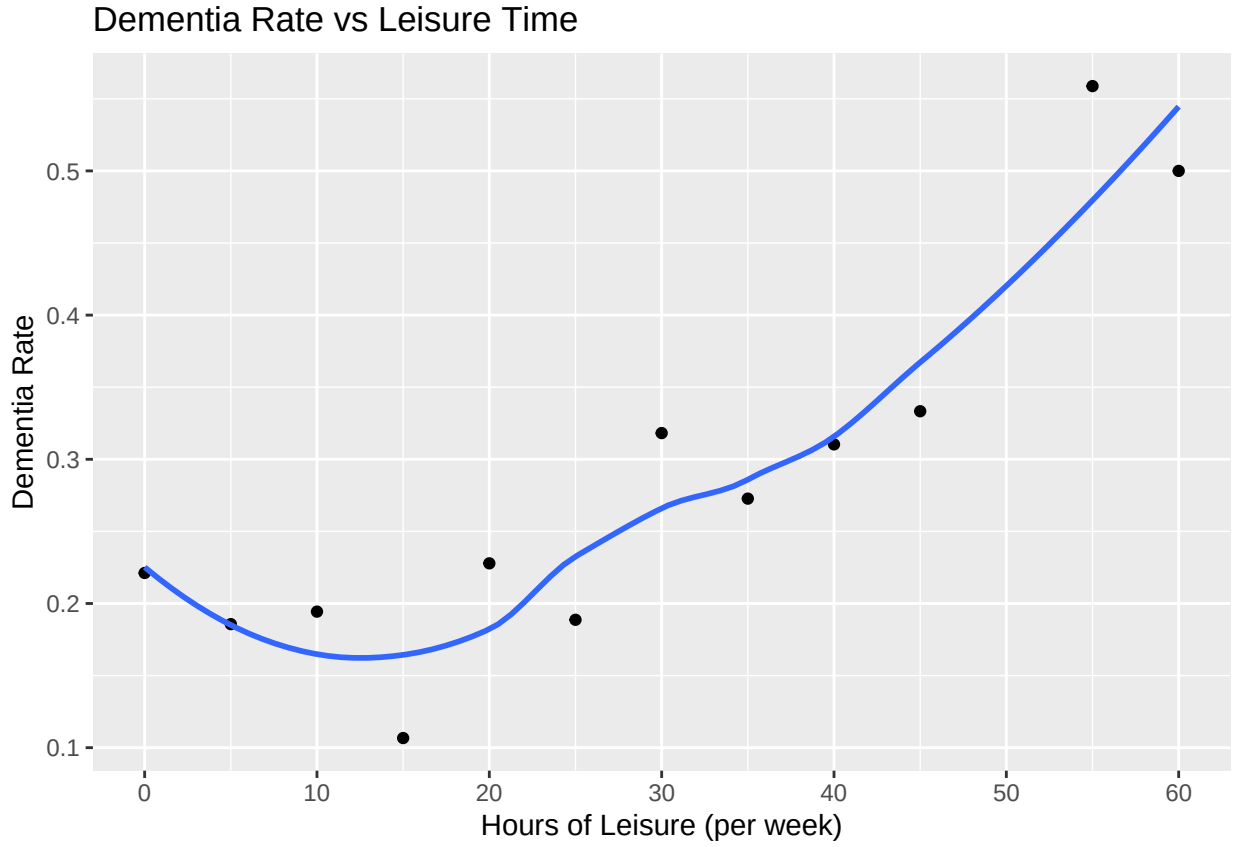
```
# create empty variable dhl to fill in for loop
dhl <- c()
hl <- c(1:13) # accompanying variable for dataframe creation

# for loop filters by segments of 5 and finds the rate in that segment before moving on
for(i in c(1:13)) {
  d6 <- filter(dcom1,
               hrs_leisure < i*5,
               hrs_leisure > (i-1)*5,
               age > 64)

  dhl <- append(dhl, sum(d6$dwtr) / length(d6$hrs_leisure))
```

```
}

# create dataframe of # vs rate
dhl_dat <- data.frame(hl, dhl)

# plot said dataframe
ggplot(dhl_dat, aes(hl, dhl)) +
  geom_point() +
  geom_smooth(se = F)  +
  scale_x_continuous(labels = c("0", "10", "20", "30", "40", "50", "60"),
                     breaks = c(1,3,5,7,9,11,13)) + # labeling appropriately
  labs(x = "Hours of Leisure (per week)", y = "Dementia Rate",
       title = "Dementia Rate vs Leisure Time")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

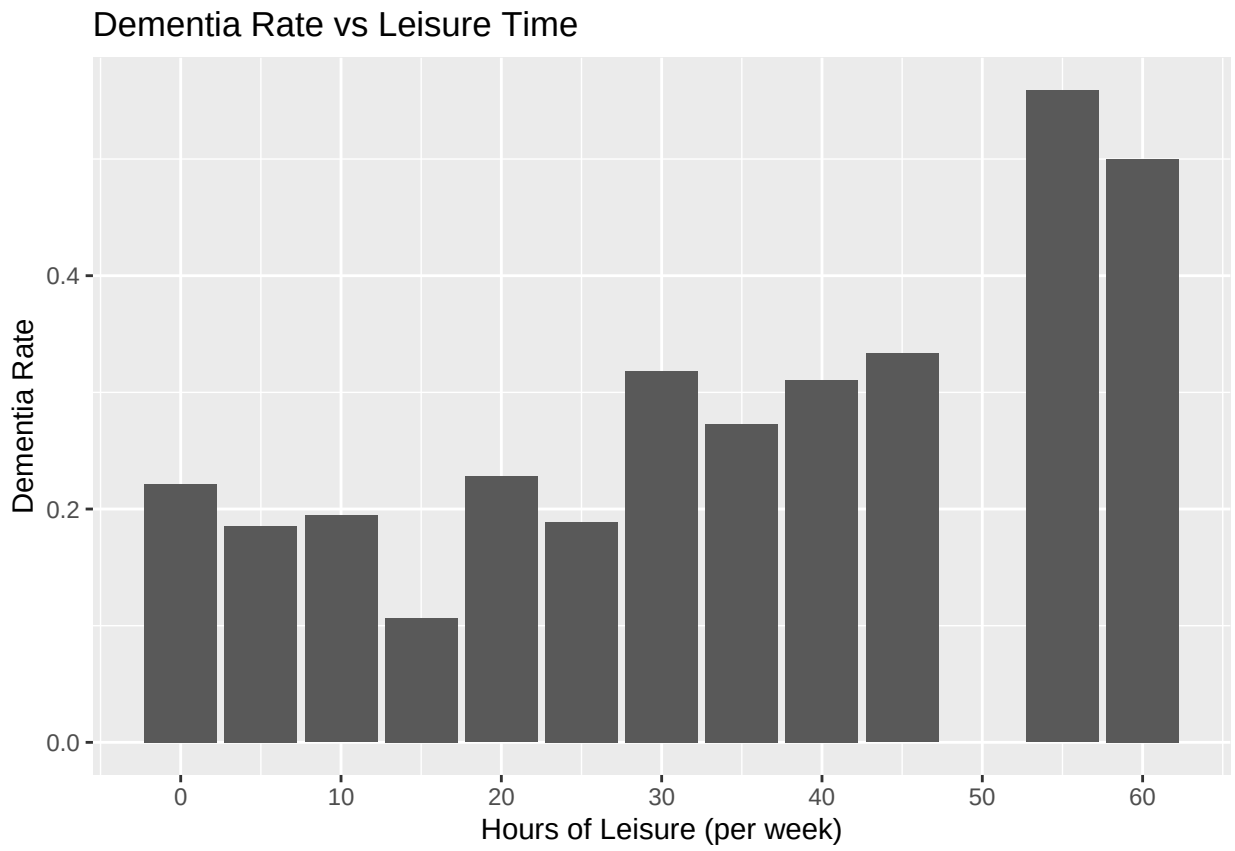
```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

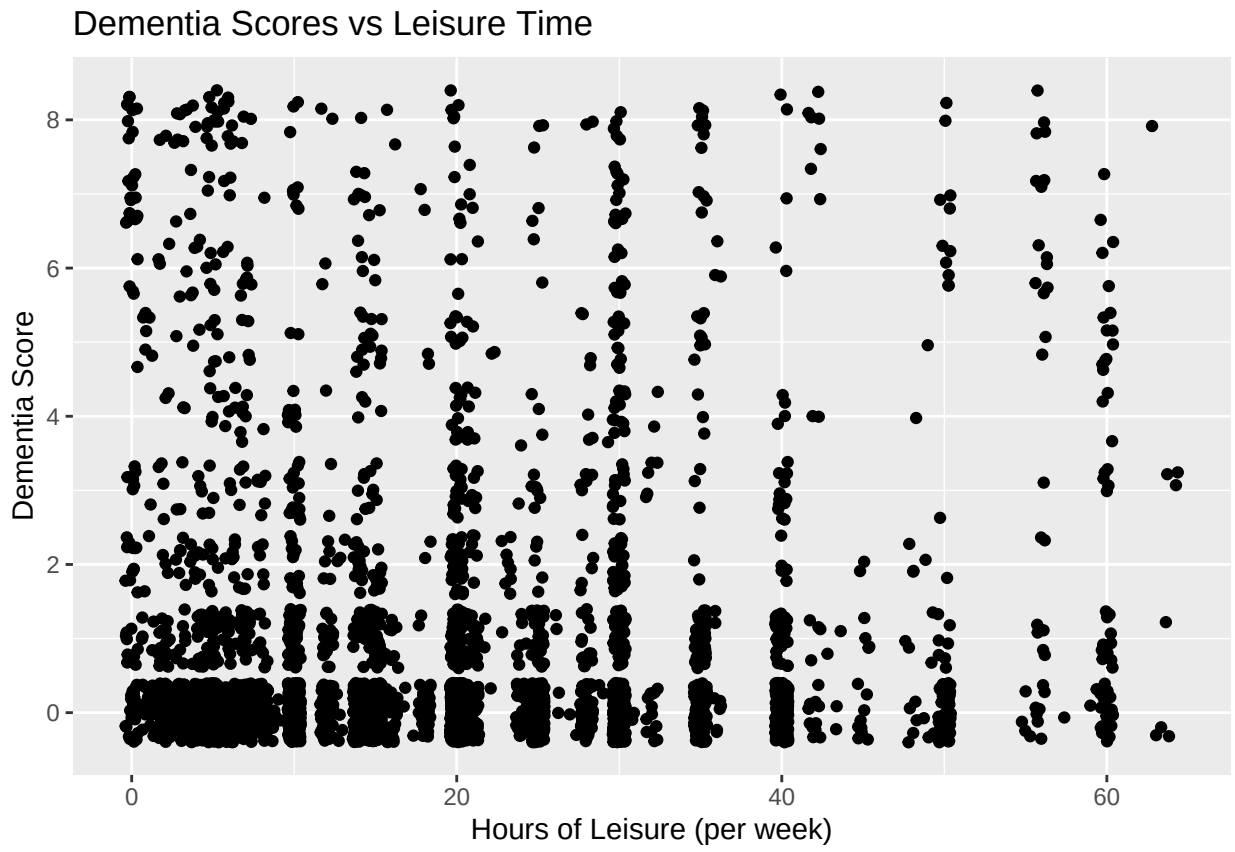



```
# plot on bar graph as well for clarity
ggplot(dhl_dat, aes(hl, dhl)) +
  geom_col() +
  scale_x_continuous(labels = c("0", "10", "20", "30", "40", "50", "60"),
                     breaks = c(1,3,5,7,9,11,13)) +
  labs(x = "Hours of Leisure (per week)", y = "Dementia Rate",
       title = "Dementia Rate vs Leisure Time")
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```

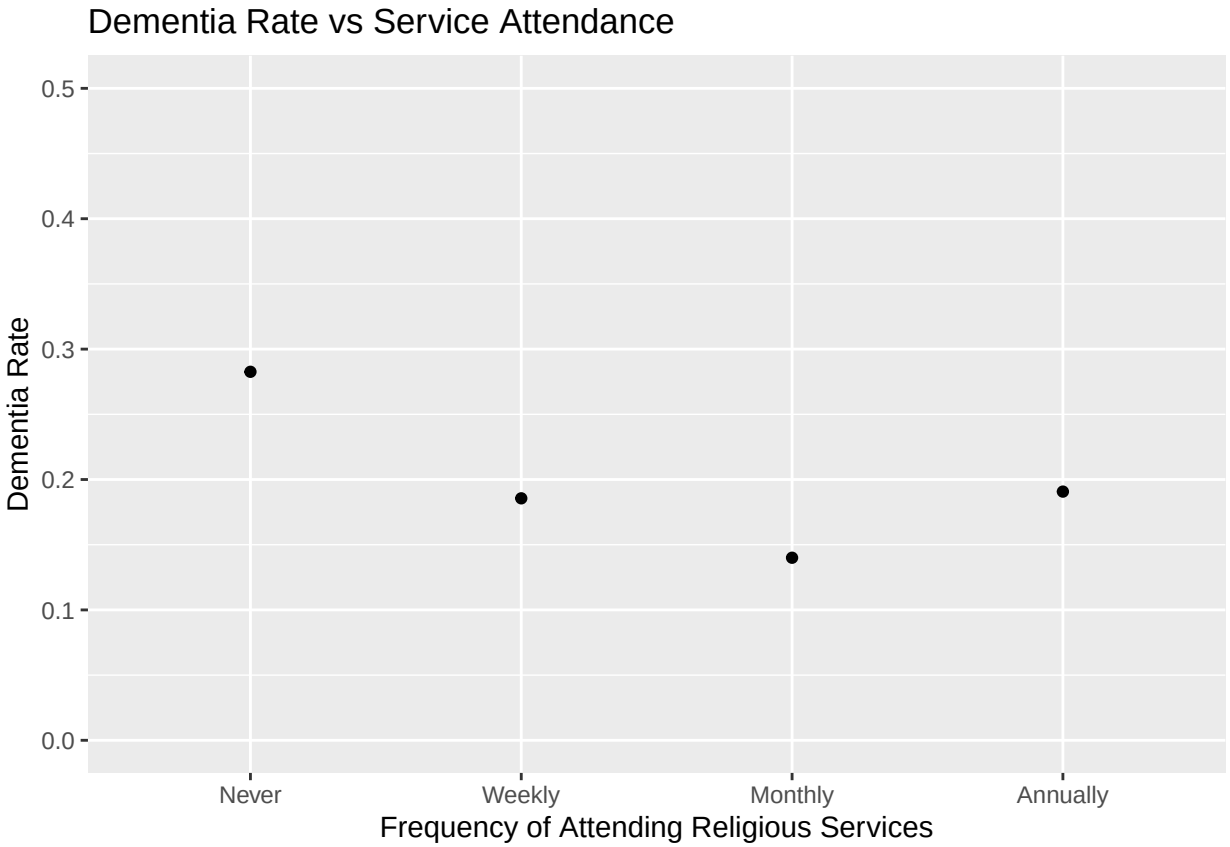



```
# standard plot similar to the ones prior
dcom1 %>%
  filter(hrs_leisure < 65, age > 64) %>%
  ggplot(aes(hrs_leisure, dscore)) +
    geom_jitter() +
    labs(x = "Hours of Leisure (per week)", y = "Dementia Score",
       title = "Dementia Scores vs Leisure Time")
```

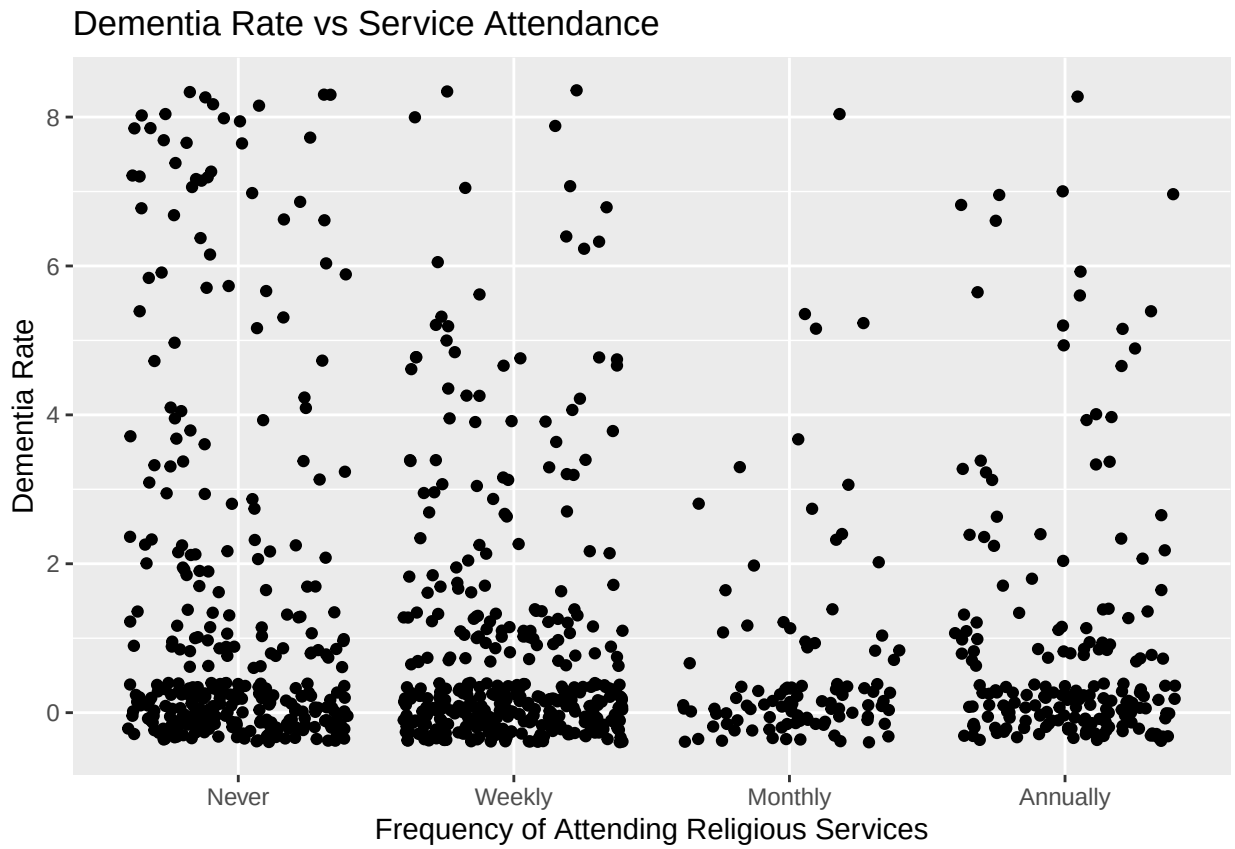


c. D vs # Services

```
dcom1$relig_svc_per <- as.character(dcom1$relig_svc_per)
# (so that graph will plot evenly across uneven numbers)

svp <- c("0","3","5","6")
# other numbers were included but with so little data as to be useless
# (e.g. daily, twice a month, etc)

k <- dcom1 %>%
  filter(age > 64, relig_svc_per == svp) %>%
  group_by(relig_svc_per) %>%
  summarize(n(), per = mean(dwtr))

ggplot(k, aes(relig_svc_per, per)) +
  geom_point() +
  y_scale +
  scale_x_discrete(labels = c("Never", "Weekly", "Monthly","Annually")) +
  labs(x = "Frequency of Attending Religious Services", y = "Dementia Rate",
       title = "Dementia Rate vs Service Attendance")
```




```
dcom1 %>%
  filter(relig_svc_per == svp, age > 64) %>%
  ggplot(aes(relig_svc_per, dscore)) +
  geom_jitter() +
  scale_x_discrete(labels = c("Never", "Weekly", "Monthly","Annually")) +
  labs(x = "Frequency of Attending Religious Services", y = "Dementia Rate",
       title = "Dementia Rate vs Service Attendance")
```




```
# average rate among religious vs non-religious individuals
dcom1 %>%
  mutate(relyn = ifelse(relig == 0, 0, 1)) %>%
  group_by(relyn) %>%
  filter(age > 64) %>%
  summarise(n(), mean(dwtr))
```

```
## # A tibble: 2 x 3
##   relyn `n()` `mean(dwtr)`
##   <dbl> <int>        <dbl>
## 1     0   244        0.152
## 2     1  3930        0.195
```

```
# average rates divided up by religion (to show a more complete picture of the variance)
dcom1 %>%
  filter(age > 64) %>%
  group_by(relig) %>%
  summarise(n(), rate = sum(dwtr) / n())
```

```
## # A tibble: 8 x 3
##   relig `n()`   rate
##   <dbl> <int>  <dbl>
## 1     0   244 0.152
## 2     1   957 0.178
## 3     2   114 0.219
## 4     8  2632 0.196
## 5    10    51 0.137
```

```
## 6    13    28 0.0714
## 7    97    15 0.267
## 8    99   133 0.301
```