

Race Against Time

Project for PSY6422, CNHN, TUOS using data from Aging, Dementia, and TBI Study of Allen Brain Map project.

Created on: 2023-05-09; Updated on: 2023-05-18

Table of Contents

I. AGING vs DEMENTIA	2
I.I Work is under-way	4
II. Allen Brain Map Databases.....	5
III. Now, let's get some LIBRARIES loaded in	5
IV. Take a swim in the DATA sea	6
IV.I. Loading data	6
IV.II. Creating tidy data.....	7
IV.III. Save tidy data	9
V. On to some VISUALIZATIONS we go.....	9
Presence of ApoE4 allele	10
Beta amyloid plaque deposition	11
Post mortem neurofibrillary tangles	12
V.IV. And then there were three: Dementia, TBI, and various markers.....	13
Brief Discussion.....	15
VI. CONCLUSION	15
VII. REFERENCES	15
RBasic.....	15
The plot thickens	15
On topic	16

Make the most of it

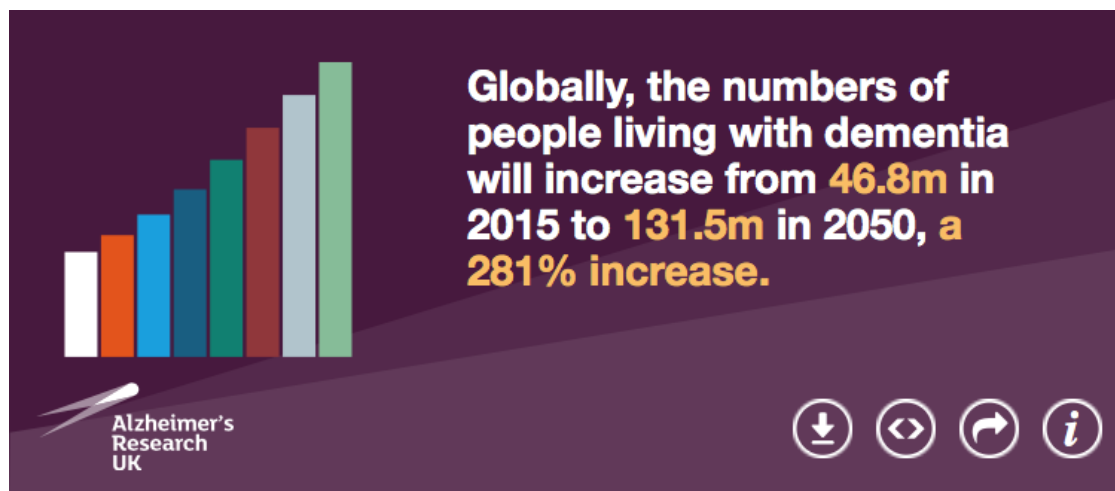
Before we begin, you can click on code in the top-right corner of the page to hide all code and simply have a read through and look at the graphs. Or dive right in. You can also hide/show individual code chunks. You can click on 'compare data on hover' at the top-right of the graphs. You can hover over the graph to find the exact count.

I. AGING vs DEMENTIA

Aging is a time-dependent process wherein the longer you live, the more your cells are degenerating, denaturing. *'Healthy aging'* is a result of this natural, inevitable deterioration; it can look different for different people. *'Clinical aging'*, however, is an accelerated degeneration of the physiological system, and often a debilitating dysfunction of the cognition.

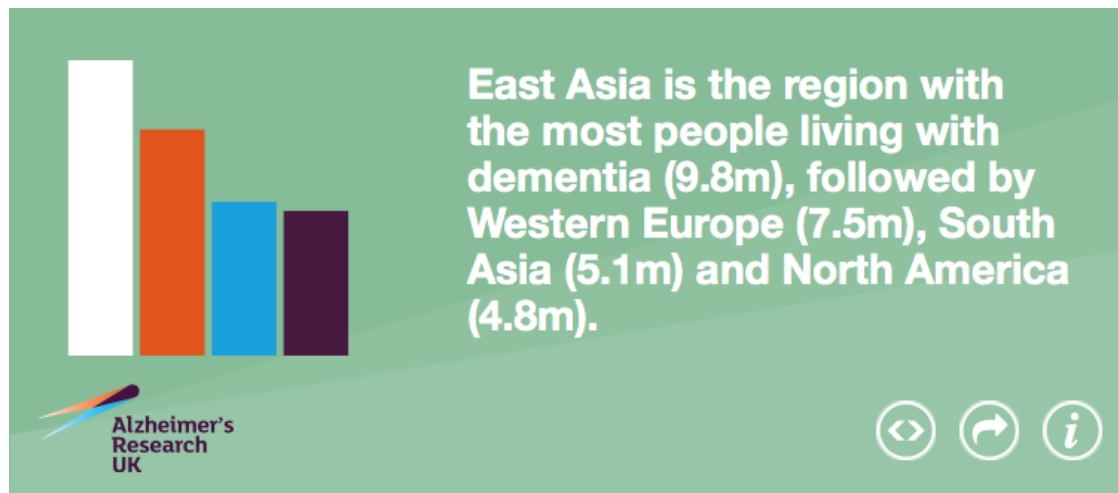
People often use aging, dementia, and Alzheimer's interchangeably. However, as mentioned above, aging can be 'healthy' or 'clinical'. **Dementia** is usually what happens when clinical aging takes place, including not exclusively memory impairment, behavioural changes, changes in motivation and emotional experience. *Alzheimer's Disease*, on the other hand, is a very specific type of dementia.

There is increasing body of research effort in the field of dementia, Alzheimer's, and other neuro-degenerative diseases; understandably so. Global prevalence is estimated at 55M people affected (2020), and this number is expected to more than double in the next thirty years.



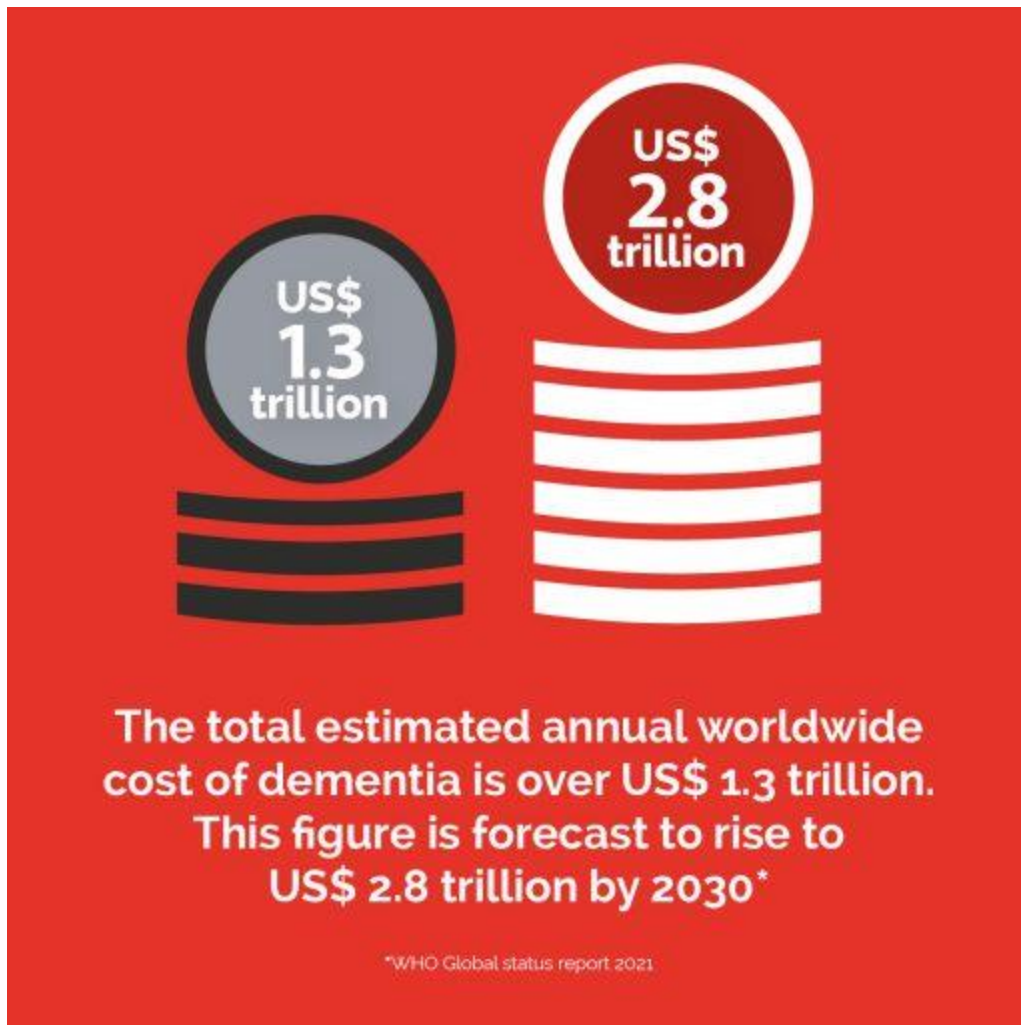
Source: *Dementia fact sheet September 2022; World Health Organisation, retrieved from dementiastatistics.org*

Of course, uber-fast changing life styles and global environment are a factor, so is the increase in the average life span and global population when it comes to correctly interpreting such numbers. Nevertheless, it is nothing short of an endemic.



Source: Prince, M et al (2015). World Alzheimer's Report 2015, The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends. Alzheimer's Disease International, retrieved from dementiastatistics.org

Dementia has psychological and physical repercussions on not only the person who experiences it but also their closest loved ones. In addition to that, chronic illnesses like this put an immense strain on the healthcare systems (that are already struggling) and the economy.



WHO Global Status Report 2021, retrieved from alzint.org

I.I Work is under-way

Clinical, research, and translational work, as well as AI integration are currently aimed at two goals:

- better management
- early detection and prevention

Both the goals rely on extensive data to

- understand disease symptomology and etymology across a varied demographic
- to identify potential risk factors and prevalent co-morbidities
- to identify potential targets for multi-pronged, early therapeutic intervention and accurate prognosis

II. Allen Brain Map Databases

The [Allen Institute](#) is an independent, non-profit organisation focusing on biosciences research. One of their projects is the [Allen Brain Map](#) is largely aimed at collecting large datasets and creating atlases based on them. In their own words, their “portal provides access to high quality data and web-based applications created for the benefit of the global research community.” These data-sets can be used by scientists, educators, and policy makers across the world.

The ABM, unsurprisingly, has a [Aging, Dementia, TBI study](#). This is where we get our [data](#) from today. The link contains data, as well as description of the data and metadata.

III. Now, let's get some LIBRARIES loaded in

The cogs of a well-oiled machine. Require regular updates though. *Can* be a bit of a pain.

Libraries Used	Purpose
tidyverse_2.0.0	transforming/wrangling data
here_1.0.1	easy file referencing/build file paths
plotly_4.10.1	create interactive graphs
dplyr_1.1.0	data manipulation
viridis_0.6.2	color-blindness friendly color palette; perceptually uniform in colour and black-and-white
htmlwidgets_1.6.2	provides a framework for creating R bindings to JavaScript libraries

```
# ----- LIBRARIES -----
```

```
# if the library is required but not installed, it will be installed first  
and then loaded
```

```
using <- function(...){ # assigning custom function to a vector; the (...) allows it to accept any number of arguments  
  libs <- unlist(list(...)) # list(...) creates a list of the arguments passed to the using function, and unlist converts it into a single vector; this is stored in libs  
  req <- unlist(lapply(libs, require, character.only=T)) # require checks if package is installed & loaded; lapply applies require to each element of libs  
  need <- libs[req==F] # extracts package names from libs that are FALSE or not loaded or installed  
  if(length(need)>0){ # checks if there are any packages in need  
    install.packages(need) # installs the packages  
    lapply(need, require, character.only=T) #loads packages  
  }  
}
```

```
# calling the functional vector; write the packages you are using here  
using("tidyverse", "here", "plotly", "dplyr", "viridis", "htmlwidgets")
```

```
# alternate code if you have only one or two packages:
# if(!require(tidyverse)){install.packages("tidyverse"); library(tidyverse)}

# you can check the version you have by using: packageVersion("viridis") for
example.
```

IV. Take a swim in the DATA sea

One of the biggest challenges is to get accessible and usable data that is not **too raw** (all those open-access fMRI scans, I'm looking at you), nor beyond your scope of understanding. A humbling experience.

The csv file used in this visualization is from the first link [here](#). Anonymity is key, so, all patients/participants had donor id and a code name. Then, there is some demographic information such as age, gender, and ethnicity. There are then quite a few clinical parameters listed, of which I was interested in

- cholesterol carrier gene allele apolipoprotein E4 (ApoE4). apo_e4_allele is the one of the biggest risk factors for late onset AD.
- amyloid deposition from 0 (none) to 3 (frequent/severe), as defined by the Consortium to Establish a Registry for Alzheimer's Disease cerad.
- the postmortem scoring of progressive AD neuropathology in different brain regions. braak stages I and II indicate "neurofibrillary tangle presence confined primarily to the transentorhinal region of the brain", stages III and IV indicate "limbic region involvement i.e. the hippocampus", and stages V and VI indicate "extensive neocortical involvement".

IV.I. Loading data

```
# ----- LOAD FILE -----

# load data file
original_df <- read_csv(here("data", "query.csv"))

# take a look at the data
# good practice, become familiar with data, columns, before beginning
head(original_df, 6)

## # A tibble: 6 × 19
##   donor_id name      age  sex  apo_e4_allele education_years
##   <dbl> <chr>    <chr> <chr> <chr>          <dbl>
## 1 326765665 H14.09.0... 87    M    N              16
## 2 326765656 H14.09.0... 95-99 M    N              17
## 3 326765654 H14.09.0... 85    M    Y              10
## 4 467056391 H15.09.1... 90-94 F    N              11
```

```

87
## 5 309335447 H14.09.0... 100+ M Y 16
0
## 6 309335457 H14.09.0... 95-99 F N 18
0
## # i 12 more variables: longest_loc_duration <chr>, cerad <dbl>,
## # num_tbi_w_loc <dbl>, dsm_iv_clinical_diagnosis <chr>, control_set
<dbl>,
## # nincds_arda_diagnosis <chr>, ever_tbi_w_loc <chr>, race <chr>,
## # hispanic <chr>, act_demented <chr>, braak <dbl>, nia_reagan <dbl>

```

One of the most crucial steps in data analysis and visualization is **data wrangling**; sometimes you wrangle the data, mostly, it wrangles you. But when it works, it's quite rewarding! This is also the step where you want to think about which variables you are interested in, what relationship could they have, and how can you accurately depict it. Pretty much the what and why.

At this stage, I have had to change the data I wanted to use, or the research question I wanted answered, or how I wanted to show it best multiple times because of my data accessibility limitations and, admittedly, my own limited coding experience. But you learn as you go.

IV.II. Creating tidy data

```

# ----- DATA WRANGLING -----

# tidy up using select to keep only relevant columns from the large original
df
selected_columns_df <- original_df %>%
  select(name, sex, apo_e4_allele, act_demented, ever_tbi_w_loc, cerad,
braak)

# check class, certain functions only work with char/numeric/factor
sapply(selected_columns_df, class)

##           name           sex apo_e4_allele  act_demented ever_tbi_w_loc
## "character"  "character"  "character"    "character"    "character"
##           cerad           braak
## "numeric"    "numeric"

# rename the columns to simplify them
renamed_df <- selected_columns_df %>%
  rename('dementia' = act_demented, 'tbi_w_loc' = ever_tbi_w_loc)

# changing character vector to numeric using gsub(); class remains unchanged!
renamed_df$dementia <- gsub("No Dementia", 0, renamed_df$dementia)
renamed_df$dementia <- gsub("Dementia", 1, renamed_df$dementia)
renamed_df$tbi_w_loc <- gsub("N", 0, renamed_df$tbi_w_loc)
renamed_df$tbi_w_loc <- gsub("Y", 2, renamed_df$tbi_w_loc)

renamed_df$dementia <- as.numeric(renamed_df$dementia) # class is changed to

```

```

numeric
renamed_df$tbi_w_loc <- as.numeric(renamed_df$tbi_w_loc)

# create new column 'disease condition' by 'adding' the numeric values on
# 'dementia' and 'TBI' columns
# I am doing this to create 4 groups that I can group the data later to for
# more meaningful comparison
wrangled_df <- renamed_df %>%
  mutate(disease_cond = renamed_df$dementia + renamed_df$tbi_w_loc)

# changing numeric back to character for easier plot making and labelling
wrangled_df$disease_cond <- as.character(wrangled_df$disease_cond)
wrangled_df$cerad <- as.character(wrangled_df$cerad)
wrangled_df$braak <- as.character(wrangled_df$braak)

# changing labels + creating new column for those labels within the df

if_df1 <- within(wrangled_df, disease_condition <- ifelse(disease_cond== "0",
  "No Dementia + No TBI with LOC",
  ifelse(disease_cond== "1", "Dementia + No TBI with
  LOC",
  ifelse(disease_cond== "2", "No Dementia +
  TBI with LOC" ,
  ifelse(disease_cond== "3", "Dementia +
  TBI with LOC", NA))))))

if_df2 <- within(if_df1, Apo_e4_allele <- ifelse(apo_e4_allele == "N", "ApoE4
allele absent",
  ifelse(apo_e4_allele ==
  "N/A", "N/A",
  ifelse(apo_e4_allele
  == "Y", "ApoE4 allele present", NA))))

if_df3 <- within(if_df2, cerad_score <- ifelse(cerad== "0", "No Aβ42
deposition",
  ifelse(cerad == "1", "Sparse
Aβ42 deposition",
  ifelse(cerad == "2",
  "Moderate Aβ42 deposition",
  ifelse(cerad ==
  "3", "Frequent Aβ42 deposition", NA))))))

if_df4 <- within(if_df3, braak_staging <- ifelse(braak == "1", "Stage 1 PM-AD
neuropath",
  ifelse(braak == "2", "Stage
2 PM-AD neuropath",
  ifelse(braak == "3",
  "Stage 3 PM-AD neuropath",
  ifelse(braak

```



```

== "4", "Stage 4 PM-AD neuropath",
ifelse(braak == "5", "Stage 5 PM-AD neuropath",
ifelse(braak == "6", "Stage 6 PM-AD neuropath", NA))))))

# remove unused column (by only including desired columns)
# rearrange columns (by simply writing the column names in desired order)
aging_df <- if_df4 %>%
  select(name, sex, disease_condition, Apo_e4_allele, cerad_score,
braak_staging)

```

There is ***probably*** a better way to handle these variables. This is what ***I could do best***.

IV.III. Save tidy data

```

# create local copy of clean df
# easier to share, especially if original file is too large
# plus better to save all the wrangling hardwork!
write_csv(aging_df, here("data", "tidy_data.csv"))

```

V. On to some VISUALIZATIONS we go

I decided to create three graphs because I felt they presented a better picture together.

V.I. Genes come first.

The apolipoprotein E (ApoE) gene is a gene that provides instructions for making the ApoE protein. The APOE gene comes in several different forms, or alleles, but the three most common ones are APOE2, APOE3, and APOE4.

Studies have found that the ApoE4 allele is associated with an increased risk of developing Alzheimer's disease and other forms of dementia. People who inherit one copy of the ApoE4 allele from one parent have an increased risk of developing Alzheimer's disease compared to people who do not have the allele. People who inherit two copies of the ApoE4 allele, one from each parent, have an even higher risk of developing Alzheimer's disease.

Of course, not everyone who has the APOE4 allele will develop dementia, and not everyone who develops dementia has the ApoE4 allele, which you will also notice in FIG.1.

V.II. House cleaning nightmare

A β 42 is a peptide that is produced by the cleavage of a larger protein called amyloid precursor protein (APP). In healthy individuals, A β 42 is cleared from the brain through various mechanisms, but in Alzheimer's disease, it accumulates in the brain, forming insoluble plaques that are toxic to neurons.

The accumulation of A β 42 is thought to be an early event in the pathogenesis of Alzheimer's disease, preceding the onset of cognitive symptoms. As A β 42 plaques accumulate, they can trigger a cascade of events that lead to neuronal damage and changes


```

# enter interaction!
dplyrplotly1 <- subdf1 %>%
  plot_ly() %>%
  add_trace(
    x= ~disease_condition, # x-axis
    y= ~count, # y-axis
    color= ~Apo_e4_allele, # grouping variable
    colors = c("ApoE4 allele absent" = '#CC1480', "N/A" = '#FF9673', "ApoE4
allele present" = '#E1C8B4'), # customize the colors
    type= 'bar', # type of plot
    name = list(),
    legendgroup = "ApoE4 Allele", # helpful to separate legends in subplot
    hovertemplate = paste(
      "<b><i>Count: %{y}</i></b><br><br>")) %>% # hover info text
  layout(hoverlabel = list( # hover info customization
    font = list(
      family = "sans-serif",
      size = 12,
      color = "black"))) %>%
  layout(xaxis = xform1, # call to vector created earlier to set x-axis
    customization
      font = "sans-serif")

# to view the fruits of your effort
# savour it
# also a good spot to check if things are working as you want them
# + play around with customizations above to see effect
dplyrplotly1

```

FIG.1: Presence of ApoE4 allele in dementia

Beta amyloid plaque deposition

```

# second plot: disease condition v CERAD score
# you know the drill

subdf2 <- aging_df %>%
  group_by(disease_condition, cerad_score) %>%
  summarise(count= n(), .groups = "drop")

xform2 <- list(categoryorder = "array",
  categoryarray = c("No Aβ42 deposition",
    "Sparse Aβ42 deposition",
    "Moderate Aβ42 deposition",
    "Frequent Aβ42 deposition"))

dplyrplotly2 <- subdf2 %>%
  plot_ly() %>%
  add_trace(x= ~cerad_score,
    y= ~count,
    color= ~disease_condition, colors = viridis_pal(option = "D")(4),

```

```

      type= 'bar',
      name = list(), legendgroup = "Disease Condition",
      hovertemplate = paste(
        "<b><i>Count: {y}</i></b><br><br>")) %>%
layout(hoverlabel = list(
  font = list(
    family = "sans-serif",
    size = 12,
    color = "black"))) %>%
layout(xaxis = xform2,
  font = "sans-serif")

```

dplyrplotly2

FIG.2: Beta amyloid plaque deposition in dementia and TBI (with loss of consciousness)

Post mortem neurofibrillary tangles

```

# third plot: disease condition v BRAAK score
# aaaand, one more time

```

```

subdf3 <- aging_df %>%
  group_by(disease_condition, braak_staging) %>%
  summarise(count= n(), .groups = "drop")

xform3 <- list(categoryorder = "array",
  categoryarray = c("Stage 1 PM-AD neuropathology",
    "Stage 2 PM-AD neuropathology",
    "Stage 3 PM-AD neuropathology",
    "Stage 4 PM-AD neuropathology",
    "Stage 5 PM-AD neuropathology",
    "Stage 6 PM-AD neuropathology"))

dplyrplotly3 <- subdf3 %>%
  plot_ly() %>%
  add_trace(x= ~braak_staging,
    y= ~count,
    color= ~disease_condition, colors = viridis_pal(option = "C")(4),
    type= 'bar',
    name = list(), legendgroup = "Disease Condition (2)",
    hovertemplate = paste(
      "<b><i>Count: {y}</i></b><br><br>")) %>%
layout(hoverlabel = list(
  font = list(
    family = "sans-serif",
    size = 12,
    color = "black"))) %>%
layout(xaxis = xform3,
  font = "sans-serif")

```

dplyrplotly3

FIG.3: Post mortem analysis of AD neuropathology pervasiveness in dementia

V.IV. And then there were three: Dementia, TBI, and various markers

United!

```
# combining the three plots into one
# may take lots of trials to get it right for your plot/ data

final <- subplot(
  dplyrplotly1 %>% layout(showlegend = TRUE),
  dplyrplotly2 %>% layout(showlegend = TRUE),
  dplyrplotly3 %>% layout(showlegend = TRUE),
  shareY = TRUE)%>% # if they share the same variable on an axis, make them
share it!
  layout(
    title = list(text = 'ApoE4 allele presence, Aβ42 deposition frequency,
and post-mortem AD neuropathology severity in dementia and TBI',
      y = 4), # title text and position
    margin = list(t = 75)) # more on position

# more ways of adding text to your plot
# here we have created a list
annotations = list(
  list(
    x = 0.2, #set the coordinates
    y = 1.0,
    text = "Disease Condition",
    xref = "paper", # how the understands the coordinate reference point
    yref = "paper",
    xanchor = "center", # more help to know where the text should go
    yanchor = "bottom",
    showarrow = FALSE # setting to true will usually point to the coordinates
specified above
  ),
  list(
    x = 0.5,
    y = 1.0,
    text = "CERAD Score",
    xref = "paper",
    yref = "paper",
    xanchor = "center",
    yanchor = "bottom",
    showarrow = FALSE
  ),
  list(
    x = 0.8,
    y = 1.0,
    text = "BRAAK Staging",
    xref = "paper",
```

```

yref = "paper",
xanchor = "center",
yanchor = "bottom",
showarrow = FALSE
),
list(
  x = 1.0,
  y = -0.15,
  text = "Source: Allen Brain Map > Aging, Dementia, TBI",
  xref = "paper",
  yref = "paper",
  showarrow = FALSE))

```

```

# one final plot making, with more instructions on where the texts go
finally <- final %>%
  layout(annotations = annotations, # refer to that list you just created
         above
         showlegend = TRUE, # legends are how we interpret the plot, usually
         legend = list(tracegroupgap = 200)) # will help separate the three
         legends created for each subplot

# grouping of legends makes the individual traces un-clickable, instead you
# interact with the whole legend
# still useful

```

Deep breaths.

```

# marvel at it
finally

```

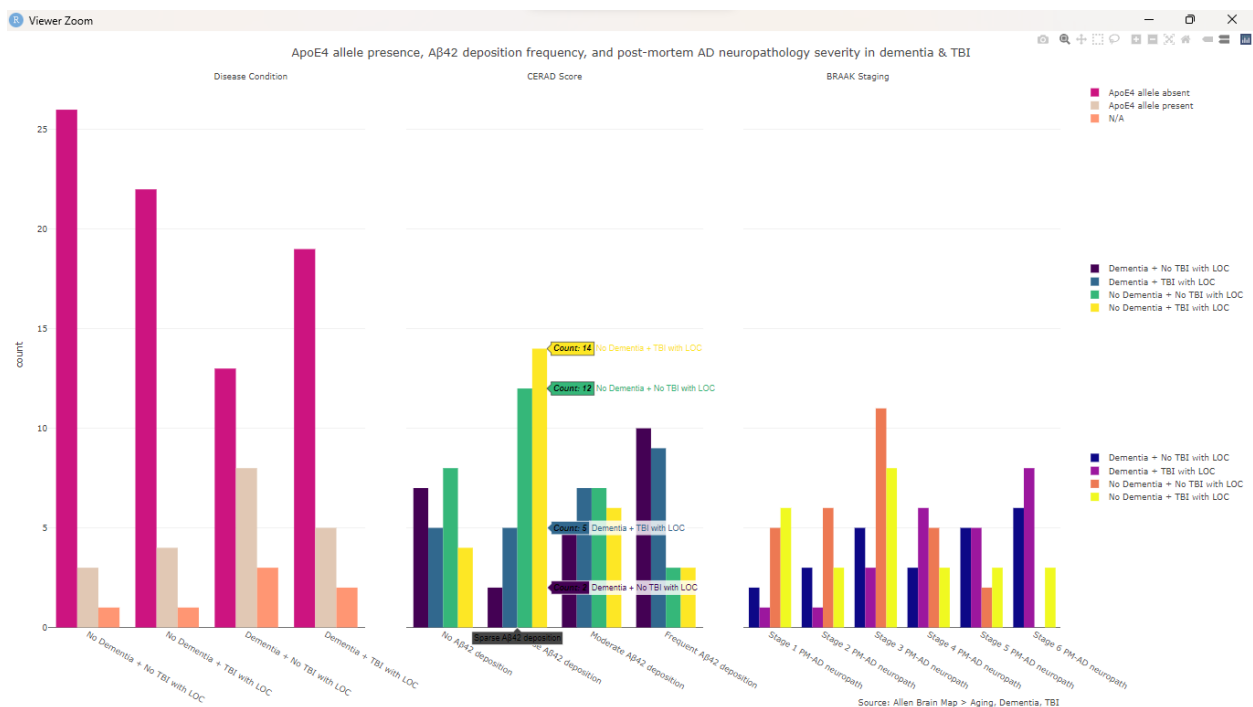


FIG.4: Dementia, TBI, and various markers

```
# save it
htmlwidgets::saveWidget(finally, file.path(fig_dir, 'Aging-Dementia-TBI-
interactive.html'))
```

Brief Discussion

Looking at these variables together shows us an interesting pattern. As expected, those with dementia generally scored higher on the various parameters. While those with no dementia or TBI tended to score the lowest. The mixed groups showed interesting results. The presence of certain bio markers, for example, did not necessarily predicted a dementia outcome. On the other hand, those without dementia but having experienced TBI may show those markers because these can have different etiologies.

Having a larger data set and running statistical analysis would be helpful in understanding these relationships and developing predictive models.

VI. CONCLUSION

As is, dementia remains a complex disease with many contributing factors. One may have all the predisposition for dementia, and still never experience it in their lifetime. On the other hand, co-morbidity such as cardiovascular disease, neuroinflammatory disease, and TBI may increase chances of developing dementia.

Well, hope that was fun and educational, maybe even morbid. That's all for now.

VII. REFERENCES

- [Acknowledging Dr. Stafford](#)
- [Acknowledging Luke Cleland](#)

RBasic

- [RMarkdown styling](#)
- [RMarkdown basic syntax](#)
- [Advanced syntax](#)
- [knitr](#)
- [ifelse](#)
- [ifthenshowme](#)
- [RMarkdown](#)
- [Table this for now](#)
- [More on Markdown](#)

The plot thickens

- [Stuff of legends](#)
- [Quit hover-ing over me](#)
- [More hover-ing](#)
- [Chart your way out](#)

- [Colour is for everyone](#)
- [Never enough legends](#)
- [More legends](#)
- [The subplot thickens](#)

On topic

- [WHO on dementia](#)
- [Alzheimer's Research UK's Dementia Statistic Hub](#)
- [Stunning visualization from Allen Institute](#)
- [Boluda, S., Toledo, J. B., Irwin, D. J., Raible, K. M., Byrne, M. D., Lee, E. B., Lee, V. M., & Trojanowski, J. Q. \(2014\). A comparison of A \$\beta\$ amyloid pathology staging systems and correlation with clinical diagnosis. *Acta neuropathologica*, 128\(4\), 543–550.](#)
- [Mufson, E. J., Malek-Ahmadi, M., Perez, S. E., & Chen, K. \(2016\). Braak staging, plaque pathology, and APOE status in elderly persons without cognitive impairment. *Neurobiology of aging*, 37, 147–153.](#)
- [Li, X., Feng, X., Sun, X., Hou, N., Han, F., & Liu, Y. \(2022\). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2019. *Frontiers in Aging Neuroscience*, 14.](#)