MGpeach

Marissa Gay

2024-03-04

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(readr)
library(dplyr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v tibble  3.1.8     v stringr 1.4.1
## v tidyr   1.2.1     v forcats 0.5.2
## v purrr   0.3.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
full <- read_csv("diagnosing_AD_data.csv")
```

```
## Rows: 174 Columns: 452
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (2): ID, class
## dbl (450): air_time1, disp_index1, gmrt_in_air1, gmrt_on_paper1, max_x_exten...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(full)
```

```
## # A tibble: 6 x 452
##   ID    air_ti~1 disp_~2 gmrt_~3 gmrt_~4 max_x~5 max_y~6 mean_~7 mean_~8 mean_~9
##   <chr>    <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 id_1      5160 1.25e-5    121.    86.9     957    6601   0.362   0.217   104.
## 2 id_2     51980 1.6 e-5    115.    83.4    1694    6998   0.273   0.145    99.4
## 3 id_3      2600 1.03e-5    230.   173.     2333    5802   0.387   0.181   201.
## 4 id_4      2130 1.03e-5    369.   183.     1756    8159   0.557   0.165   276.
```

```
## 5 id_5      2310 6.86e-6    258.   111.      987    4732   0.266   0.145   185.
## 6 id_6      1920 1.14e-5    200.   110.     1548    6260   0.213   0.143   155.
## # ... with 442 more variables: mean_jerk_in_air1 <dbl>,
## #   mean_jerk_on_paper1 <dbl>, mean_speed_in_air1 <dbl>,
## #   mean_speed_on_paper1 <dbl>, num_of_pendown1 <dbl>, paper_time1 <dbl>,
## #   pressure_mean1 <dbl>, pressure_var1 <dbl>, total_time1 <dbl>,
## #   air_time2 <dbl>, disp_index2 <dbl>, gmrt_in_air2 <dbl>,
## #   gmrt_on_paper2 <dbl>, max_x_extension2 <dbl>, max_y_extension2 <dbl>,
## #   mean_acc_in_air2 <dbl>, mean_acc_on_paper2 <dbl>, mean_gmrt2 <dbl>, ...
```

```
full %>% group_by(class) %>%
  summarise(mean(air_time1),
            mean(air_time2),
            mean(air_time3),
            mean(air_time4),
            mean(air_time5),
            mean(air_time6),
            mean(air_time7),
            mean(air_time8),
            mean(air_time9),
            mean(air_time10),
            mean(air_time11),
            mean(air_time12),
            mean(air_time13),
            mean(air_time14),
            mean(air_time15),
            mean(air_time16),
            mean(air_time17),
            mean(air_time18),
            mean(air_time19),
            mean(air_time20),
            mean(air_time21),
            mean(air_time22),
            mean(air_time23),
            mean(air_time24),
            mean(air_time25))
```

```
## # A tibble: 2 x 26
##   class mean(a~1 mean(~2 mean(~3 mean(~4 mean(~5 mean(~6 mean(~7 mean(~8 mean(~9
##   <chr>    <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 H        4782.   2860.   1793.   2648.   1491.   4457.  14943.   1773.   1885.
## 2 P        6507.  10277.   4951.  13211.   5977.  12895.  12169.   6088.   7895.
## # ... with 16 more variables: `mean(air_time10)` <dbl>,
## #   `mean(air_time11)` <dbl>, `mean(air_time12)` <dbl>,
## #   `mean(air_time13)` <dbl>, `mean(air_time14)` <dbl>,
## #   `mean(air_time15)` <dbl>, `mean(air_time16)` <dbl>,
## #   `mean(air_time17)` <dbl>, `mean(air_time18)` <dbl>,
## #   `mean(air_time19)` <dbl>, `mean(air_time20)` <dbl>,
## #   `mean(air_time21)` <dbl>, `mean(air_time22)` <dbl>, ...
```

```
AD_airtime <- ggplot() +
  geom_point(aes(x = ))
```

Including Plots

You can also embed plots, for example:

```
a=1
b=6
c=15
access <- data.frame(matrix(nrow=174))

for (i in 1:25) {
  access <- access %>%
    mutate(full[a:(a+2)], full[b:(b+1)], full[c:(c+1)])

  a <- a+18
  b <- b+18
  c <- c+18

}

access <- access %>%
  mutate(full[451:452])

access <- subset(access, select = -(matrix.nrow...174.))
```

```
d <- 2

#for task 25

for (i in (1:25)) {
  df25 <- data.frame(access[1], access[d:(d+6)])
  df25 <- df25 %>%
    mutate(access[177])

  assign( paste("Task", i, sep = "_"), df)
  d <- d+7
}

head(df25)
```

```
##     ID air_time25 disp_index25 max_x_extension25 max_y_extension25
## 1 id_1     104485     4.95e-05             10066             13235
## 2 id_2     171940     6.98e-05              7365             15282
## 3 id_3      33545     5.60e-05              7688             14127
## 4 id_4     113275     5.79e-05              6397             14913
## 5 id_5      35290     4.25e-05              4624             15532
## 6 id_6      30815     4.99e-05              4023             15335
##   num_of_pendown25 paper_time25 total_time25 class
## 1               71        40120       144605     P
## 2              129       126700       298640     P
## 3               74        45480        79025     P
## 4              123        67945       181220     P
## 5               92        37285        72575     P
## 6               76        43790        74605     P
```

```r
#task23subset

d <- 2


for (i in (1:23)) {
  df23 <- data.frame(access[1], access[d:(d+6)])
  df23 <- df23 %>%
    mutate(access[177])

  assign( paste("Task", i, sep = "_"), df)
  d <- d+7
}

head(df23)
```

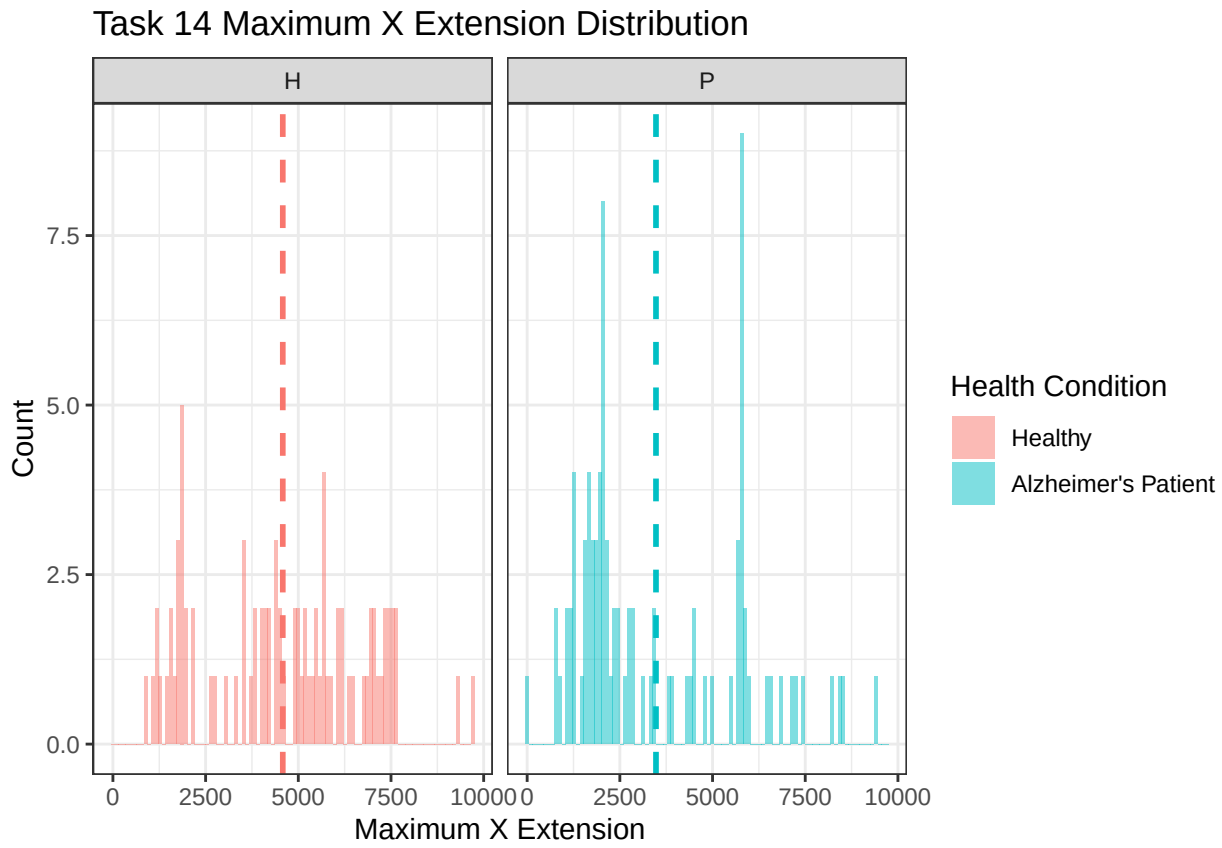
```
##     ID air_time23 disp_index23 max_x_extension23 max_y_extension23
## 1 id_1      10965     9.48e-06               788               5828
## 2 id_2      14660     1.08e-05               848               5800
## 3 id_3       7330     1.09e-05              1338               8208
## 4 id_4       7205     1.03e-05              1429               6663
## 5 id_5       5340     7.55e-06               761               5183
## 6 id_6       4485     8.26e-06               556               4955
##   num_of_pendown23 paper_time23 total_time23 class
## 1               12         5195        16160     P
## 2               14        15240        29900     P
## 3               12         6535        13865     P
## 4               11         6380        13585     P
## 5               12         4805        10145     P
## 6               14         6240        10725     P
```

```r
#task14subset

d <- 2


for (i in (1:14)) {
  df14 <- data.frame(access[1], access[d:(d+6)])
  df14 <- df14 %>%
    mutate(access[177])

  assign( paste("Task", i, sep = "_"), df)
  d <- d+7
}

head(df14)
```

```
##     ID air_time14 disp_index14 max_x_extension14 max_y_extension14
## 1 id_1       2910     8.02e-06              1670               2945
## 2 id_2      61980     3.44e-05              3467              11755
## 3 id_3       4675     9.78e-06              2109               6584
## 4 id_4       6795     1.43e-05              1281              10056
## 5 id_5       3605     1.10e-05               786               8071
```

```
## 6 id_6       4495     1.52e-05              1225              9723
##   num_of_pendown14 paper_time14 total_time14 class
## 1                5         6680         9590     P
## 2               30        41695       103675     P
## 3                4         5250         9925     P
## 4               16        10485        17280     P
## 5               15         7360        10965     P
## 6               13         9980        14475     P
```

```
mean_data <- aggregate(max_x_extension14 ~ class, data = df14, FUN = mean)

ggplot() + theme_bw()+
  geom_histogram(aes(x = max_x_extension14, fill = class),
                           data = df14,
                           bins = 100,
                           alpha = 0.5)+
  facet_wrap(~class)+
  scale_fill_discrete(name="Health Condition",
                        breaks=c("H", "P"),
                        labels=c("Healthy", "Alzheimer's Patient"))+
   geom_vline(data = mean_data, aes(xintercept = max_x_extension14, color = class),
              linetype = "dashed", size = 1, show.legend = FALSE)+
  labs(x= "Maximum X Extension", y = "Count")+
  ggtitle("Task 14 Maximum X Extension Distribution" )
```

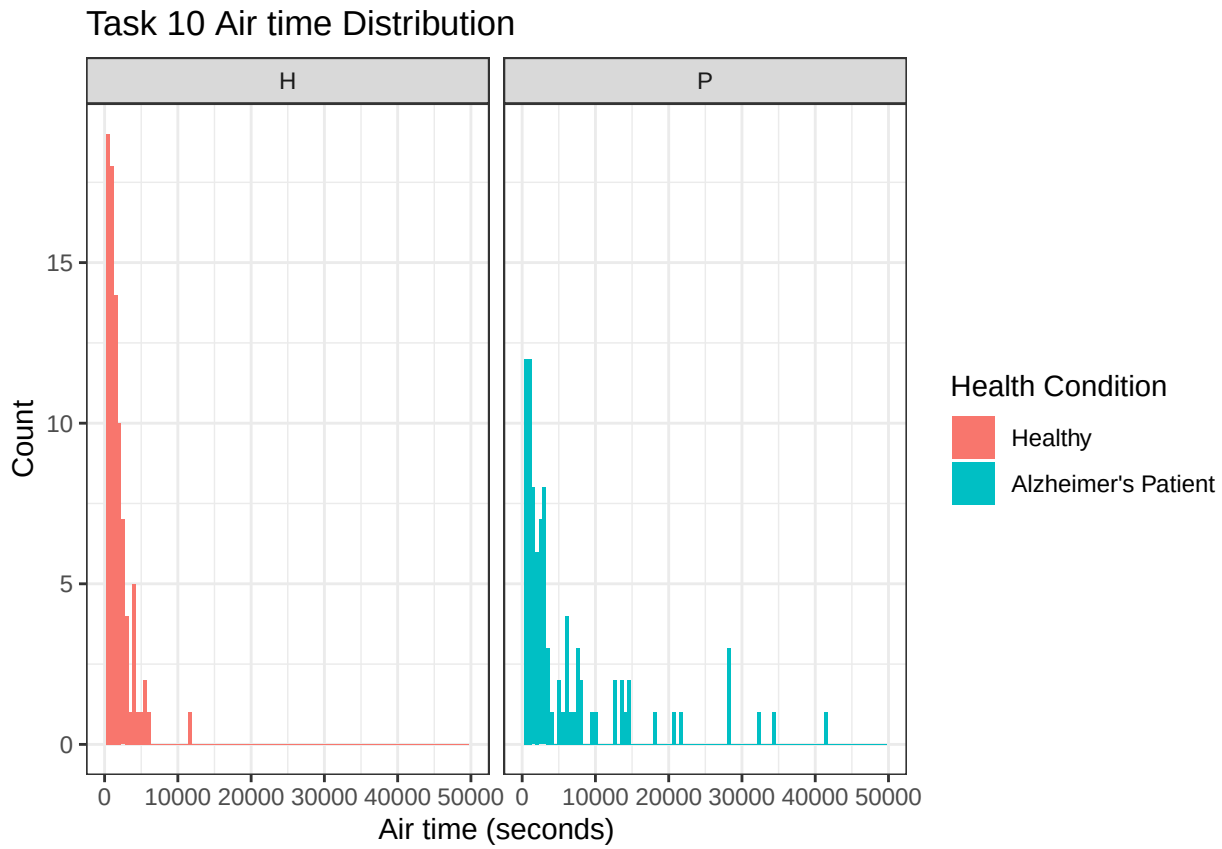


```r
#task14subset

d <- 2


for (i in (1:10)) {
  df10 <- data.frame(access[1], access[d:(d+6)])
  df10 <- df10 %>%
    mutate(access[177])

  assign( paste("Task", i, sep = "_"), df)
  d <- d+7
}

head(df10)
```

```
##     ID air_time10 disp_index10 max_x_extension10 max_y_extension10
## 1 id_1       2575     5.66e-06              1499              3210
## 2 id_2       5995     8.20e-06              1158              2644
## 3 id_3        505     5.48e-06              1779              2096
## 4 id_4      20775     4.50e-06               580              2238
## 5 id_5        800     4.54e-06               887              1762
## 6 id_6        945     7.94e-06              1366              2651
##   num_of_pendown10 paper_time10 total_time10 class
## 1                5         3300         5875     P
## 2                8         9280        15275     P
## 3                2         3135         3640     P
## 4                7         4475        25250     P
## 5                5         3190         3990     P
## 6                3         5590         6535     P
```

```r
ggplot() + theme_bw()+
  geom_histogram(aes(x = air_time10, fill = class),
                          data = df10,
                          bins = 100)+
  facet_wrap(~class)+
  xlim(0,50000)+
  labs(x= "Air time (seconds)",
       y = "Count",
       fill = "Health Condition")+
   scale_fill_discrete(name="Health Condition",
                          breaks=c("H", "P"),
                          labels=c("Healthy", "Alzheimer's Patient"))+
  ggtitle("Task 10 Air time Distribution")
```

```
## Warning: Removed 4 rows containing missing values (geom_bar).
```

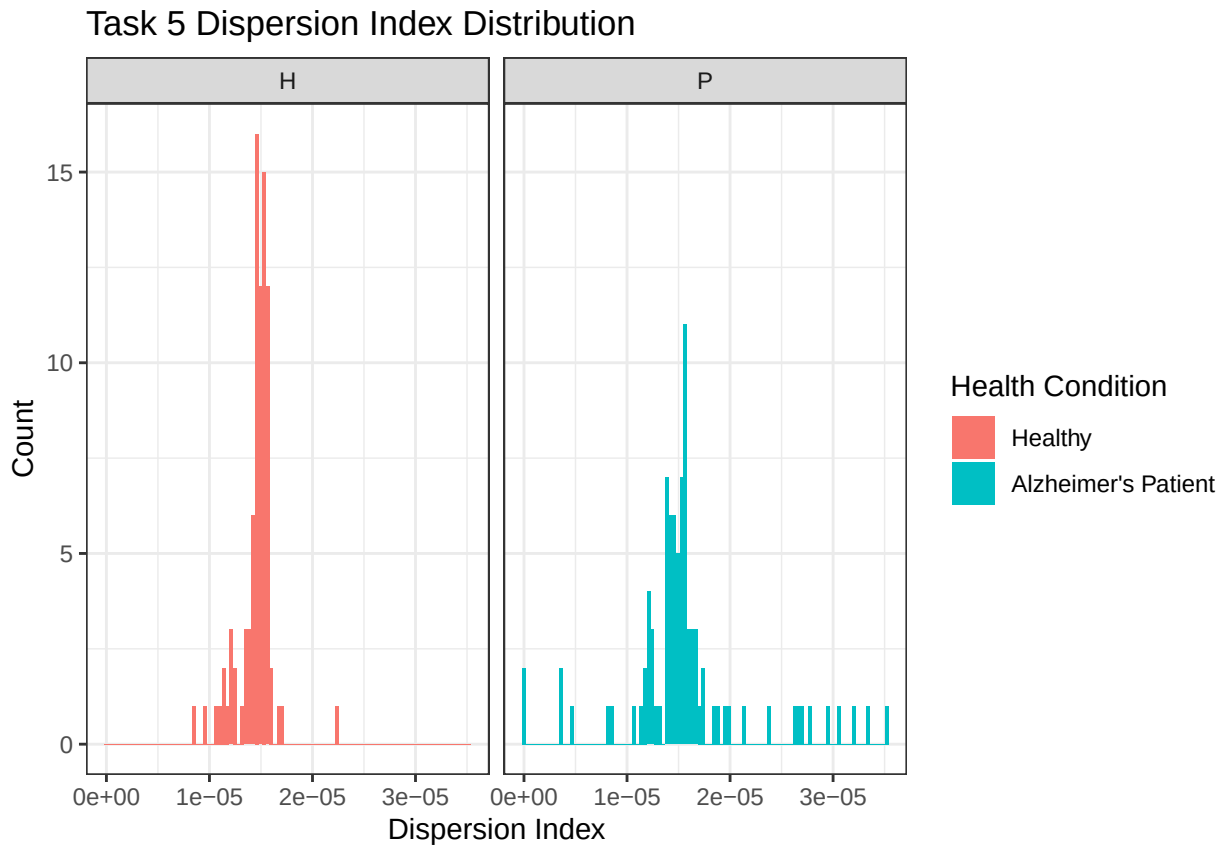



```
d <- 2


for (i in (1:5)) {
  df5 <- data.frame(access[1], access[d:(d+6)])
  df5 <- df5 %>%
    mutate(access[177])

  assign( paste("Task", i, sep = "_"), df)
  d <- d+7
}

head(df5)
```

```
##     ID air_time5 disp_index5 max_x_extension5 max_y_extension5 num_of_pendown5
## 1 id_1     66034    1.55e-05            10933             3651              17
## 2 id_2     12875    3.36e-05             5667             5503               8
## 3 id_3       680    1.43e-05             2556             2245               1
## 4 id_4     10735    1.55e-05             2535             2426              14
## 5 id_5      1050    3.50e-06             2394             1194               1
## 6 id_6      2400    1.57e-05             2486             2447               8
##   paper_time5 total_time5 class
## 1       64885      130919     P
## 2       31055       43930     P
## 3       11630       12310     P
## 4       41200       51935     P
```

```
## 5        3300        4350     P
## 6       35570       37970     P
```

```
ggplot() + theme_bw()+
  geom_histogram(aes(x = disp_index5, fill = class),
                           data = df5,
                           bins = 100)+
  facet_wrap(~class)+
  labs(x= "Dispersion Index",
       y = "Count",
       fill = "Health Condition")+
   scale_fill_discrete(name="Health Condition",
                       breaks=c("H", "P"),
                       labels=c("Healthy", "Alzheimer's Patient"))+
  ggtitle("Task 5 Dispersion Index Distribution")
```



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.