STAD94 REPORT

Joanna Yijing Rong
University of Toronto Scarborough
2016 Summer

Data Analysis of a Research Study - The Impact of an Interdisciplinary, Case-Managed Diabetes Team on Diabetes Self Management, Diabetes-Related Hospitalizations and Emergency Department Visits

TABLE OF CONTENTS

PART I - INTRODUCTION

BACKGROUND & PURPOSE OF THE STUDY

With the imposing health impact of diabetes on Canadians and the existing substantial financial burden, the Ministry of Health and Long-Term Care launched the Centres for Complex Diabetes Care (CCDC). CCDC was developed to offer a time-limited intervention for individuals whose diabetes and related conditions require a more intensive and comprehensive team-based approach.Once individuals are better able to manage their diabetes and diabetes-related complications, they are discharged and referred to a community-based program for ongoing support, if required. However, there is limited research on the impact of CCDCs' case management model on healthcare utilization and health outcomes. Therefore, the purpose of this study is to evaluate the impact of an interdisciplinary, case-managed diabetes team on patients' usage of emergency departments and hospital admissions, health outcomes, and Quality of Life Index.

HYPOTHESIS

The hypothesis of the study is that an interdisciplinary, case-managed diabetes team will have a positive impact on patients' usage of emergency departments and hospital admissions, health outcomes, and Quality of Life Index.

METHODOLOGY

This study used information gathered on patients who attended the CCDC and is based on retrospective chart review on all patients discharged from the CCDC on or before June 30, 2015.

DATA ANALYSIS

All data is analyzed with RStudio statistical analysis software. Comparison between continuous variables will use t-tests. Chi-square tests will be used to compare categorical variables. Analysis of Variance (ANOVA) will be used to test differences between means. Additional analyses may be required depending on results.

PART II - FUNCTIONS

This project relies on many well-structured functions, which allow us to reuse it again and again once they are defined. These functions have made it easier for us to process a large amount of data quickly and efficiently. In this section of the report, I will discuss what each function is and their purposes.

"EXTRACTMORBIDITY""

This function produces "True" or "False" base on whether or not each patient has code for a particular type of comorbidity, chronic morbidity or mental health diagnosis. The purpose of defining this function is to allow us to count the total numbers of patients with each type of comorbidity, chronic morbidity and mental health diagnosis.

"EXTRACTCO"&"EXTRACTCHRONIC"&"EXTRACTMH"

These three functions all serve the same purposes: they utilize "extractmorbidity" to label different types of comorbidity, chronic morbidity and mental health diagnosis. For example, if a patient has code "8"(8 is the code for Neuropathy) in his/her comorbidity columns, "extractco" produces "True" for Neuropathy. Similarly, if that patient does not have code "8", " extractco" produces "False" for Neuropathy.

"HASREPEAT""

The purpose of this function is to help us find which patient has duplicated comorbidity codes. Coupling with the apply function, "hasrepear" can scan along the specific comorbidity, and produce the corresponding code number if a patient has more than one of the same comorbidity code; likewise, the function produces "NA" if the patient has only one comorbidity code. For example, if a patient has more than one "1" in his/her comorbidity column, "hasrepear" will produce "1" as a result.

"COUNT23""

The role of this function is to capture how many times "23" (the code for mental health) appear in each patient's comorbidity column. "count23" produce numbers that represent numbers of time that "23" appear in each patient's comorbidity column. Also, using "which" function with "count23", we were able to find out precisely which patients have more than one of "23" in their comorbidity columns.

"COUNTMH"

Similarly, this function is designed to count the total numbers of actual mental health diagnosis each patient has in the mental health diagnosis columns. For example, if a patient has three different types of mental health diagnosis, "countmh" produces 3 as a result.

"TIDYDDSCODE""

The purpose of this function is to select only the non-numerical values (ie., 999, 99, "." and blank) from each DDS column.

"TIDYDDSNUMBERS"

In contrast, this function selects only numbers (below 20) from each DDS columns.

PART III - DATA ORGANIZATION

Data organization is the most important component, because only well-organized data can allow us to accurately assess, replicate, and evaluate any research results. However, with such large piece of data, organizing and cleaning them up was the most challenging and time-consuming part. In this section, I will highlight some of the most prominent problems that we encountered as well as how we fixed them.

PROBLEM #1

One of the areas we intended to investigate was to see whether certain conditions (ie., comorbidity, chronic morbidity and Mental Health diagnosis) have any impacts on patients' health outcomes before and after the treatments. Specifically, there are 23 types of comorbidity, 16 types of chronic morbidity and 12 types of mental health diagnosis in total, and each type is coded with a number.

SOLUTION FOR PROBLEM #1

To approach the question, we first decided to count the total numbers of patients with each type of comorbidity, chronic morbidity and mental health diagnosis. To do so: we began with creating a function called "extractmorbidity", which produces "True" or "False" base on whether the patients have or do not have the code for a particular type of comorbidity, chronic morbidity or mental health diagnosis. We also created three other functions called "extractco", "extractchronic" and "extractMH", which utilizes "extractmorbidity" to name various types of comorbidity, chronic morbidity and mental health diagnosis. As a result, each patient has their own list of different types of comorbidity, chronic morbidity or mental health diagnosis, with "Ture" or "False" displaced underneath each of them- this indicates if a patient is with or without a specific condition. For example, if a patient has code "3" ("3" is the code for Hypertension) in his/her comorbidity column, "extractco" produces "True" for Hypertension. Similarly, if that patient does not have code "3", " extractco" produces "False" for Hypertension. Finally, with the help of "gather" and "subset", we were finally able to obtain the total numbers of patients for each type of comorbidity, chronic-morbidity and mental health diagnosis.

PROBLEM#2

We found that Excel and Rstudio produced different numbers of patients with certain types of comorbidity. Specifically, the differences are found in the following types of comorbidity: other, obesity(BMI>30), Hyperlipidemia, cardiovascular disease, hypertension, Mental health. For example, Excel produced 335 patients with "other" types of comorbidity, while Rstudio only produced 230 of them.
To investigate the problem, we built a function called "hasrepeat" to help us spot precisely which patient has

duplicated comobidity codes. We were surprised to find out that about 80 patients have repeated codes for the same type of co-morbidity. For example, patient#2 has two "1" (codes for "other"); patient#312 has two "11" (codes for "Obesity (BMI > 30)). This was very puzzling to me: if patient A has three "13" (code fo "Hyperlipidemia") in his/her comorbidity section, does each "13" codes for a different kind/level of Hyperlipidemia? or are they all representing the same thing - Hyperlipidemia?

SOLUTION FOR PROBLEM #2

I pointed out this issue to the principal investigator immediately. With couple times of back and forth email communications, we came to realized that the repeated comorbidity codes are due to raw data input errors. Eventually, most of the duplicate comorbidities for anything that does not make sense (hypertension, obesity, hyperlipidemia, etc) have been corrected. But there are two exclusions: if patients had "Other" selected more than once, we have decided to leave those as duplicates given the fact that these are each separate issues that don't have their own code. As well, there is one patient whose comorbidity is repeated because he is both blind and deaf. We decided to waive that one as it is a minor issue which might not have any significant effects on data analysis.

PROBLEM #3

In contrast with other types of comorbidity, repeated codes for one particular type of comorbidity - "mental health" is a different story. Each "23" (code for mental health) represents a particular type Mental Health diagnosis. For example, if a patient has code "23" five times in his comorbidity section, he/she should have five different types of mental health diagnosis under his mental health diagnosis section. However, we have to ensure that the number of mental health diagnosis is consistent with the number of repeated mental health codes.

SOLUTION FOR PROBLEM#3

To make sure that the number of mental health diagnosis is the same as the number of repeated mental health codes, we solved this by employing a "subtraction" method. First, we created two functions called "count23" coupled with the apply function to capture numbers of "23" (code for mental health) each patient has in the comorbidity columns. Similarly, we created another function called "countmh" coupled with the apply function to count for total numbers of actual mental health diagnosis each patient has in their mental health diagnosis columns. Lastly, using the table function, we were able to find that there were 96 patients who have inconsistent numbers of mental health diagnosis compared with numbers of repeated mental health codes! After checking with the principle investigator, it turns out the person who helped add that new mental health diagnosis data missed about 80 diagnoses! At the end, the missing diagnosis data have been updated.

PROBLEM #4

While we were analyzing the distance in km from the patient's home to the center (very useful information to understand patient catchment), we found that there is one huge outlier shown on the Halton site boxplot. That outlier indicated a patient who travels over 750km to get to Halton location. This outlier very questionable because the average travel distance was about 10 km, and the maximum travel distances(other than the 750km) is about 100km.

SOLUTION FOR #4

After pointing it out to the principal investigator, we immediately found out that the patient travels 7.5km instead of 750km. That suspicious outlier is a result of error. We also learned that boxplot is very useful in detecting outliers and potential errors. It is a reliable tool for both data analysis and data organization.

PROBLEM #5

Another area that we are interested in looking at is the changes in QoL index (which is DDS – Diabetes Distress Survey). There are four different types of diabetes distress: emotional burden, physician-related distress, Regimen-related distress and interpersonal distress. Each patient has their DDS score recorded before and after the treatment. However, when we produced some boxplots to compare the changes of DDS score before and after treatment, we noticed many questionable outliers that don't make sense (ie., some patient have DDS score 999 while the median score is about 1.3). We soon realized that there are many unexpected values (ie., 999, 99, "." and blank) in the raw data. Originally, 999 indicates "missing values", but in Rstudio, Boxplot takes 999 as "normal" numbers, which is what led to the strange outliers. Thus we need to come up with a solution to exclude those "non-number", and only use normal number values for the analysis.

SOLUTION TO PROBLEM#5

To solve this problem, we created two functions to separate the non-numerical values from the numbers, and both of them utilize "as.character" and "as.numeric" techniques. One of the function is called "TidyDDScode" which picks up only the non-numerical values (i.e.,., 999, 99, "." and blank) from each DDS column. Another function called "TidyDDSnumbers" picks up only numbers (below 20) from each DDS columns. Then, we redefined each type of DDS column using those two functions, which are included in the final data frame. For example, there are two forms of DDS emotional burden index: "DDS.Emotional.burden.Enrollment" include only numbers, and "DDS.Emotional.burden.Enrollment.code" include only non-numberical values. This way, we can conduct boxplot or other analysis using DDS index that is only consisted of numbers.

DATA ORGANIZATION CODES

```
#'Turn list of co/chronic mobidity and mental health into "ture" and "false"
#'
#' @param x the levels for patients in the original data frame to be turned i
nto "ture" and "false"
#' @param possible numbers of possible levels for co/chronic mobidity and men
tal health
#' @return vectors of "ture" and "false" according to the levels of co/chroni
c mobidity and mental health
extractmorbidity=function(x,possible) #possible=#of possible values for 16 or
 23 values for chronic / co-mobidity
{
  if(x[1]==999||is.na(x[1]))
  {
    return(rep(F,possible))
  }else{
```

```
    has=rep(F,possible)
    for(i in x){
      has[i]=T
    }
    return(has)
  }
}

hasrepeat=function(x,possible)
{
  tab=table(x)
  m=max(tab)
  w=names(which.max(tab))
  ifelse(m>1,w,NA)
}

#x=c(2,3,2,3,4)
#table(x)
#tab=table(t(data2[2,14:23]))
#tab["23"]

#'built a function that will count numbers of 23(mental health in comorbidity
) in one patient
count23=function(x)
{
  tab=table(t(x))
  ans=ifelse(is.na(tab["23"]),0,tab["23"])
}

#count.discharged.status=function(x)
#{
  #tab=table(t(x))
  #ans=ifelse(is.na(tab["2"]),0,tab["2"])
#}

#'built a function that count the total numbers of actual mental health issue
 for each patient
countmh=function(x)
{
  return(sum(table(t(x))))
}
```

```
#made fuction that label the specific values of each co-mobidity
extractco=function(x)
{
  thenames=c("Other","Eating Disorders","Hypertension","Cardiovascular diseas
e","Chronic kidney disease",
             "Retinopathy or Other eye disease"," Non-healing wounds (greater
 than 3 months)","Neuropathy","Liver disease (fatty liver)",
             "Peripheral vascular disease","Obesity (BMI > 30)","Current mali
gnancy/cancer treatment","Hyperlipidemia",
             "Thyroid disease (other endocrinopathies POCT, Cushings)","Demen
tia","Pulmonary disease (COPD, Asthma)","Obstructive sleep apnea",
             "HIV/AIDS","Pancreas Diseases","Celiac Disease","Genetic Syndrom
es","Deafness and/or Blindness","Mental health")
  M=extractmorbidity(x,23)
  names(M)=thenames
  return(M)

}

#made functionthat label the specific values for each chronic-morbidity
extractchronic=function(x)
{
  thenames=c("Other(Chronic)","Neuropathy","Retinopathy","Blindness","Cardiov
ascular disease(Chronic)","Wounds (non-healing)",
             "Amputation","Skin conditions (cutaneous manifestations)","Lipoh
ypertrophy","Hypoglycemia unawareness",
             "Diabetic myonecrosis","Foot problems (Charcot's)","Stiff man's
syndrome","Hearing impairment","Fractures","Nephropathy")
  M=extractmorbidity(x,16)
  names(M)=thenames
  return(M)
}

extractMH=function(x)
{
  thenames=c("Depressive Disorder","Obsessive-compulsive and related disorder
s","Schizophrenia spectrum and other psychotic disorders",
             "Substance-related and addictive disorders","Bipolar Disorders",
"Anxiety disorders",
             "Neurodevelopmental disorder","Trauma and Stress-related disorde
rs","Personality Disorders","Feeding and Eating Disorders",
             "Sleep-wake disorders","Type Unspecified")

  possible=12
```

```
  if(is.na(x[1]))
  {
    y=rep(F,possible)
  }else{
    has=rep(F,possible)
    for(i in x){
      if(!is.na(i)){
        has[i]=T
      }
    }
    y=has
  }
  names(y)=thenames
  return(y)
}



#'function that capture non numeric values
TidyDDScode=function(x){
  y=as.character(x)
  z=as.numeric(y)
  code=ifelse(is.na(z),y,ifelse(z<20,"numeric",y))
  return(code)
}

#'function that capture numeric values that are below 20
TidyDDSnumbers=function(x){
  y=as.character(x)
  z=as.numeric(y)
  code=ifelse(is.na(z),NA,ifelse(z<20,z,NA))
  return(code)
}



# read data in from csv
#
# a paragraph with more details about function
#
# @param file.name name of imported file
#
# @return dataframe containing all variables as appropriatly dates, factor an
d logicals
```

```
# @export
# @import dplyr
# @import lubridate
suppressMessages(library(dplyr))

## Warning: package 'dplyr' was built under R version 3.1.2

Readdata=function(file.name)
{
  data=read.csv("inst/extdata/20160715.csv",header=T)
  #refer the column names by numbers

  #take the rows that are not NA for location and gender
  #cut = converted the numbers into factors --> converted the numbers to fact
ors of the following variables
  #library(dplyr)
  data %>%
    dplyr::slice(1:622) ->data2


  locationf=cut(data2[,2],breaks=c(0,1.2,3),labels=c("HALTON","MISSISSAUGA"))


  genderf=cut(data2[,4],breaks=c(0,1.5,3),labels=c("M","F"))

  agef=cut(data2[,3],breaks=c(17,45,65,80,95))

  repeatreferral=cut(data2[,7],breaks=c(-1,0.5,2),labels=c("No","Yes"))


  referralfrom=cut(data2[,8],breaks=c(-1,1.5,2.5,3.5,4.5,6),labels=c("primary
","self","specialist","hospital","DEC"))


  socialissue=cut(data2[,9],breaks=c(0,1.5,2.5,3.5,4.5,5.5,6.5,7.5,8.5,9.5,11
,1000),
                  labels=c("social","community","income","elderly","smoke","d
rugs","housing","mobility","language","education","none"))
  #consider changing the short names to the actual names

  profcareplan=cut(data2[,64],breaks=c(-1,0.5,2,1000),labels=c("No","Yes","N/
A"))


  SelfMGoals=cut(data2[,65],breaks=c(-1,0.5,2,1000),labels=c("No","Yes","N/A"
```

```
))


  ExtentSelfM=cut(data2[,66],breaks=c(-1,0.5,1.5,2.5,3.5,1000),labels=c("None
","Some","Most","All","N/A"))


  Pdischwithtrans=cut(data2[,67],breaks=c(-1,0.5,2,1000),labels=c("No","Yes",
"N/A"))


  DischSt=cut(data2[,69],breaks=c(0,1.5,2.5,3.5),labels=c("withdrawal","disch
arged","death"))


  Endorefer=cut(data2[,87],breaks=c(-1,0.5,2),labels=c("no","yes"))


  #extract the dates and truned them into Rdate
  #obtain different duration between admit date and discharge date

  #library(lubridate)
  referdate=lubridate::dmy(data2[,6])


  admitdate=lubridate::dmy(data2[,34])


  dischargedate=lubridate::dmy(data2[,68])



  #dis - data2[,73] #column 73 is numbers of weeeks between admitdate and dis
charge date as calculated in spreadsheet
  #summary(as.numeric(dis) - data2[,73])  #checking if the calculation is con
sisting
  #column Bu- Time in CCDC doesn't appear in the final spread sheet

  #co-mo is from column 14-23
  #chronic comlication is from column 24-33

  #made the function for "T=pts has both morbidity and F=pts doesn't have mor
bidity"
```

```
  como=t(apply(data2[,14:23],1,extractco))
  comorep=apply(data2[,14:23],1,hasrepeat)
  comorep
  which(comorep=="3")
  comomhrep=apply(data2[,14:23],1,count23)
  comomhrep #'counted numbers of "23" for each patient among the comorbidity
columns --> fixed: no patients has repeated 23 any more!

  which(comomhrep>1) #' no patients are found that has more than one 23
  actualmhrep=apply(data2[,74:78],1,countmh)
  actualmhrep #'counted numbers of actual mental health issue for each patien
t among the mental health sections
  table(comomhrep>=1,actualmhrep>=1)
  which(comomhrep>=1 & actualmhrep==0) # not an issue any more--> pt354 doesn
't have any mental healthy diagnosis listed, but has 23 coded in the comorbid
ity section

  table(comorep) #show which coded comorbidity has repeates --> only code "1"
 is repeated, but agree on ignoring it (also row61 has more 22 twice, but it'
s all good! )
  which(!is.na(comorep)) #shows which patients has more than one unique code
  #only co-morbidity has a lot of repeats

  chronic=t(apply(data2[,24:33],1,extractchronic))
  chronicrep=apply(data2[,24:33],1,hasrepeat)
  chronicrep
  table(chronicrep)
  #no repeats are found in chronic complications

  MH=t(apply(data2[,74:78],1,extractMH))
  head(MH)
  MHrep=apply(data2[,74:78],1,hasrepeat)
  MHrep
  table(MHrep)
  #no repeats found for MH diagnosis


  #make a new dataframe
  como=as.data.frame(como)
  chronic=as.data.frame(chronic)
  MH=as.data.frame(MH)
  dataframe=data.frame(locationf,age=data2[,3],genderf,agef,distance=data2[,5
],referdate,repeatreferral,referralfrom,socialissue,
                       DDS.Emotional.burden.Enrollment=TidyDDSnumbers(data2[,
```

```
10]),DDS.Emotional.burden.Enrollment.code=TidyDDScode(data2[,10]),
                      DDS.Physician.related.distress.Enrollment=TidyDDSnumbe
rs(data2[,11]),DDS.Physician.related.distress.Enrollment.code=TidyDDScode(dat
a2[,11]),
                      DDS.Regimen.related.distress.Enrollment=TidyDDSnumbers
(data2[,12]),DDS.Regimen.related.distress.Enrollment.code=TidyDDScode(data2[,
12]),
                      DDS.Interpersonal.distress.Enrollment=TidyDDSnumbers(d
ata2[,13]),DDS.Interpersonal.distress.Enrollment.code=TidyDDScode(data2[,13])
,
                      como,chronic,admitdate,data2[,35:39],admit.waist=as.nu
meric(as.character(data2[,40])),data2[,41:49],
                      DCweight=as.numeric(as.character(data2[,50])),
                      data2[,51:58],
                      DDS.Emotional.burden.Discharge=TidyDDSnumbers(data2[,5
9]),DDS.Emotional.burden.Discharge.code=TidyDDScode(data2[,59]),
                      DDS.Physician.related.distress.Discharge=TidyDDSnumber
s(data2[,60]),DDS.Physician.related.distress.Discharge.code=TidyDDScode(data2
[,60]),
                      DDS.Regimen.related.distress.Discharge=TidyDDSnumbers(
data2[,61]),DDS.Regimen.related.distress.Discharge.code=TidyDDScode(data2[,61
]),
                      DDS.Interpersonal.distress.Discharge=TidyDDSnumbers(da
ta2[,62]),DDS.Interpersonal.distress.Discharge.code=TidyDDScode(data2[,62]),
                      data2[,63],profcareplan,
                      SelfMGoals,ExtentSelfM,Pdischwithtrans,dischargedate,D
ischSt,data2[,70:73],MH,data2[,79:86],Endorefer)
  return(dataframe)
}

dataframe=suppressWarnings(Readdata())
head(dataframe)

##   locationf age genderf    agef distance  referdate repeatreferral
## 1    HALTON  73       M (65,80]        6 2013-05-07             No
## 2    HALTON  60       F (45,65]       16 2013-05-10             No
## 3    HALTON  68       M (65,80]       18 2013-05-17             No
## 4    HALTON  66       F (65,80]       16 2013-04-29             No
## 5    HALTON  69       F (65,80]        6 2013-05-13             No
## 6    HALTON  67       M (65,80]        1 2013-05-21             No
##   referralfrom socialissue DDS.Emotional.burden.Enrollment
## 1      primary        none                             1.2
## 2      primary      social                             3.0
## 3      primary        none                             1.2
```

```
## 4      primary        none                            1.2
## 5      primary        none                            2.4
## 6         self        none                            4.6
##   DDS.Emotional.burden.Enrollment.code
## 1                              numeric
## 2                              numeric
## 3                              numeric
## 4                              numeric
## 5                              numeric
## 6                              numeric
##   DDS.Physician.related.distress.Enrollment
## 1                                       1.3
## 2                                       2.0
## 3                                       1.5
## 4                                       1.3
## 5                                       1.3
## 6                                       2.5
##   DDS.Physician.related.distress.Enrollment.code
## 1                                        numeric
## 2                                        numeric
## 3                                        numeric
## 4                                        numeric
## 5                                        numeric
## 6                                        numeric
##   DDS.Regimen.related.distress.Enrollment
## 1                                     1.4
## 2                                     4.0
## 3                                     3.2
## 4                                     1.4
## 5                                     1.8
## 6                                     5.6
##   DDS.Regimen.related.distress.Enrollment.code
## 1                                     numeric
## 2                                     numeric
## 3                                     numeric
## 4                                     numeric
## 5                                     numeric
## 6                                     numeric
##   DDS.Interpersonal.distress.Enrollment
## 1                                   1.0
## 2                                   3.0
## 3                                   1.0
## 4                                   1.0
```

```
## 5                                    1.7
## 6                                    4.7
##   DDS.Interpersonal.distress.Enrollment.code Other Eating.Disorders
## 1                                    numeric  TRUE            FALSE
## 2                                    numeric  TRUE            FALSE
## 3                                    numeric FALSE            FALSE
## 4                                    numeric  TRUE            FALSE
## 5                                    numeric FALSE            FALSE
## 6                                    numeric FALSE            FALSE
##   Hypertension Cardiovascular.disease Chronic.kidney.disease
## 1         TRUE                  FALSE                   TRUE
## 2         TRUE                  FALSE                  FALSE
## 3         TRUE                  FALSE                  FALSE
## 4         TRUE                  FALSE                  FALSE
## 5         TRUE                  FALSE                  FALSE
## 6         TRUE                   TRUE                  FALSE
##   Retinopathy.or.Other.eye.disease
## 1                            FALSE
## 2                            FALSE
## 3                            FALSE
## 4                            FALSE
## 5                            FALSE
## 6                            FALSE
##   X.Non.healing.wounds..greater.than.3.months. Neuropathy
## 1                                        FALSE      FALSE
## 2                                        FALSE      FALSE
## 3                                        FALSE      FALSE
## 4                                        FALSE      FALSE
## 5                                        FALSE      FALSE
## 6                                        FALSE      FALSE
##   Liver.disease..fatty.liver. Peripheral.vascular.disease
## 1                       FALSE                       FALSE
## 2                        TRUE                       FALSE
## 3                       FALSE                       FALSE
## 4                       FALSE                       FALSE
## 5                       FALSE                       FALSE
## 6                       FALSE                       FALSE
##   Obesity..BMI...30. Current.malignancy.cancer.treatment Hyperlipidemia
## 1               TRUE                              FALSE           TRUE
## 2               TRUE                              FALSE           TRUE
## 3              FALSE                              FALSE          FALSE
## 4               TRUE                              FALSE           TRUE
## 5              FALSE                              FALSE           TRUE
```

```
## 6              FALSE                                    FALSE          FALSE
##   Thyroid.disease..other.endocrinopathies.POCT..Cushings. Dementia
## 1                                                  FALSE    FALSE
## 2                                                  FALSE    FALSE
## 3                                                  FALSE    FALSE
## 4                                                  FALSE    FALSE
## 5                                                  FALSE    FALSE
## 6                                                  FALSE    FALSE
##   Pulmonary.disease..COPD..Asthma. Obstructive.sleep.apnea HIV.AIDS
## 1                            FALSE                  FALSE    FALSE
## 2                            FALSE                  FALSE    FALSE
## 3                            FALSE                  FALSE    FALSE
## 4                            FALSE                  FALSE    FALSE
## 5                            FALSE                  FALSE    FALSE
## 6                            FALSE                  FALSE    FALSE
##   Pancreas.Diseases Celiac.Disease Genetic.Syndromes
## 1             FALSE          FALSE             FALSE
## 2             FALSE          FALSE             FALSE
## 3             FALSE          FALSE             FALSE
## 4             FALSE          FALSE             FALSE
## 5             FALSE          FALSE             FALSE
## 6             FALSE          FALSE             FALSE
##   Deafness.and.or.Blindness Mental.health Other.Chronic. Neuropathy.1
## 1                     FALSE         FALSE          FALSE        FALSE
## 2                     FALSE          TRUE          FALSE         TRUE
## 3                     FALSE         FALSE           TRUE        FALSE
## 4                     FALSE         FALSE          FALSE        FALSE
## 5                     FALSE         FALSE          FALSE         TRUE
## 6                     FALSE         FALSE          FALSE        FALSE
##   Retinopathy Blindness Cardiovascular.disease.Chronic.
## 1       FALSE     FALSE                           FALSE
## 2       FALSE     FALSE                           FALSE
## 3        TRUE     FALSE                            TRUE
## 4       FALSE     FALSE                           FALSE
## 5        TRUE     FALSE                            TRUE
## 6       FALSE     FALSE                           FALSE
##   Wounds..non.healing. Amputation
## 1                FALSE      FALSE
## 2                FALSE      FALSE
## 3                FALSE      FALSE
## 4                FALSE      FALSE
## 5                FALSE      FALSE
## 6                FALSE      FALSE
```

```
##   Skin.conditions..cutaneous.manifestations. Lipohypertrophy
## 1                                      FALSE           FALSE
## 2                                      FALSE           FALSE
## 3                                      FALSE           FALSE
## 4                                      FALSE           FALSE
## 5                                      FALSE           FALSE
## 6                                      FALSE           FALSE
##   Hypoglycemia.unawareness Diabetic.myonecrosis Foot.problems..Charcot.s.
## 1                    FALSE                FALSE                     FALSE
## 2                    FALSE                FALSE                     FALSE
## 3                     TRUE                FALSE                     FALSE
## 4                    FALSE                FALSE                     FALSE
## 5                    FALSE                FALSE                     FALSE
## 6                    FALSE                FALSE                     FALSE
##   Stiff.man.s.syndrome Hearing.impairment Fractures Nephropathy  admitdate
## 1                FALSE              FALSE     FALSE       FALSE 2013-05-10
## 2                FALSE              FALSE     FALSE       FALSE 2013-05-17
## 3                FALSE              FALSE     FALSE        TRUE 2013-05-28
## 4                FALSE              FALSE     FALSE       FALSE 2013-05-21
## 5                FALSE              FALSE     FALSE        TRUE 2013-05-21
## 6                FALSE              FALSE     FALSE       FALSE 2013-05-23
##   admit.A1C admit.Blood.Pressure..Systolic.
## 1       9.8                             143
## 2      12.2                             150
## 3       8.3                             120
## 4       8.8                             126
## 5      10.3                             140
## 6       8.6                             170
##   admit.Blood.Pressure..Diastolic. admit.Weight..kg. admit.BMI admit.waist
## 1                               74              97.0      31.6         112
## 2                               90              98.0      38.0         132
## 3                               58              90.3      27.7         109
## 4                               68              91.8      33.0         120
## 5                               93              52.4      23.0          91
## 6                               88              79.0      27.4         113
##   admit.HDL.Cholesterol admit.LDL.Cholesterol admit.Triglycerides
## 1                  1.03                  1.54                2.09
## 2                  1.26                  2.68                1.18
## 3                  0.91                  1.17                0.55
## 4                  1.21                  2.08                0.88
## 5                  1.95                  2.65                0.79
## 6                  1.35                  1.24                1.27
##   admit.Micro.Albumin admit.ACR admit.eGFR D.C.A1C
```

```
## 1               373.7      29.9          56      NA
## 2                14.1       0.8          87     6.6
## 3                29.0       4.0          76     8.0
## 4                66.0       4.5          90     7.6
## 5                87.3      18.6          15     9.4
## 6                33.0       4.1          80     8.0
##   D.C.Blood.Pressure..Systolic. D.C.Blood.Pressure..Diastolic. DCweight
## 1                            NA                             NA       NA
## 2                           142                             70    101.6
## 3                           128                             60     89.6
## 4                           110                             78     96.6
## 5                           110                             60     53.6
## 6                           120                             84     80.3
##   D.C.BMI D.C.Waist.circumference..cm. D.C.HDL.Cholesterol
## 1      NA                           NA                  NA
## 2    39.6                        132.0                1.24
## 3    27.5                        110.0                1.09
## 4    36.1                        114.3                1.19
## 5    23.8                         78.0                2.09
## 6    27.7                         97.0                1.33
##   D.C.LDL.Cholesterol D.C.Triglycerides D.C.Micro.Albumin D.C.ACR D.C.eGFR
## 1                  NA                NA                NA      NA       NA
## 2                2.18              0.98               9.0     0.5       94
## 3                1.28              0.44               7.0     0.8       79
## 4                2.01              0.79              57.0     5.8       96
## 5                1.81              0.77              90.7    34.9       15
## 6                1.31              1.63              18.0     2.9       75
##   DDS.Emotional.burden.Discharge DDS.Emotional.burden.Discharge.code
## 1                             NA                                <NA>
## 2                            2.0                             numeric
## 3                            1.0                             numeric
## 4                            1.2                             numeric
## 5                             NA                                 999
## 6                             NA                                <NA>
##   DDS.Physician.related.distress.Discharge
## 1                                       NA
## 2                                      1.0
## 3                                      1.0
## 4                                      1.5
## 5                                       NA
## 6                                       NA
##   DDS.Physician.related.distress.Discharge.code
## 1                                          <NA>
```

```
## 2                                      numeric
## 3                                      numeric
## 4                                      numeric
## 5                                          999
## 6                                        <NA>
##   DDS.Regimen.related.distress.Discharge
## 1                                     NA
## 2                                      1
## 3                                      2
## 4                                      1
## 5                                     NA
## 6                                     NA
##   DDS.Regimen.related.distress.Discharge.code
## 1                                        <NA>
## 2                                     numeric
## 3                                     numeric
## 4                                     numeric
## 5                                        <NA>
## 6                                        <NA>
##   DDS.Interpersonal.distress.Discharge
## 1                                   NA
## 2                                    1
## 3                                    1
## 4                                    1
## 5                                   NA
## 6                                   NA
##   DDS.Interpersonal.distress.Discharge.code data2...63. profcareplan
## 1                                      <NA>          NA          Yes
## 2                                   numeric        3.10          Yes
## 3                                   numeric        3.90          Yes
## 4                                   numeric        3.65          Yes
## 5                                      <NA>          NA          Yes
## 6                                      <NA>          NA          Yes
##   SelfMGoals ExtentSelfM Pdischwithtrans dischargedate    DischSt
## 1        Yes        None             N/A    2013-11-06 withdrawal
## 2        Yes        Some             Yes    2014-04-30 discharged
## 3        Yes        Most             Yes    2014-06-17 discharged
## 4        Yes        Most             Yes    2014-08-15 discharged
## 5        Yes        Some             N/A    2015-03-20      death
## 6        Yes        Some             Yes    2013-12-12 discharged
##   X..of.Diabetes.Related.ER.Visits.in.year.prior.to.CCDC.registration
## 1                                                                   0
## 2                                                                   1
```

```
## 3                                                                    3
## 4                                                                    0
## 5                                                                    1
## 6                                                                    0
##   X..of.Diabetes.Related.Hospital.visits.in.year.prior.to.CCDC.registratio
n
## 1
0
## 2
0
## 3
2
## 4
0
## 5
3
## 6
2
##   D.C.Total...of.Diabetes.Related.ER.Visits
## 1                                        0
## 2                                        0
## 3                                        0
## 4                                        0
## 5                                        7
## 6                                        1
##   D.C.Total...of.Diabetes.Related.Hospital.Visits Depressive.Disorder
## 1                                               0               FALSE
## 2                                               0                TRUE
## 3                                               0               FALSE
## 4                                               0               FALSE
## 5                                               2               FALSE
## 6                                               0               FALSE
##   Obsessive.compulsive.and.related.disorders
## 1                                      FALSE
## 2                                      FALSE
## 3                                      FALSE
## 4                                      FALSE
## 5                                      FALSE
## 6                                      FALSE
##   Schizophrenia.spectrum.and.other.psychotic.disorders
## 1                                                FALSE
## 2                                                FALSE
## 3                                                FALSE
```

```
## 4                                                 FALSE
## 5                                                 FALSE
## 6                                                 FALSE
##   Substance.related.and.addictive.disorders Bipolar.Disorders
## 1                                     FALSE             FALSE
## 2                                     FALSE             FALSE
## 3                                     FALSE             FALSE
## 4                                     FALSE             FALSE
## 5                                     FALSE             FALSE
## 6                                     FALSE             FALSE
##   Anxiety.disorders Neurodevelopmental.disorder
## 1             FALSE                       FALSE
## 2             FALSE                       FALSE
## 3             FALSE                       FALSE
## 4             FALSE                       FALSE
## 5             FALSE                       FALSE
## 6             FALSE                       FALSE
##   Trauma.and.Stress.related.disorders Personality.Disorders
## 1                               FALSE                FALSE
## 2                               FALSE                FALSE
## 3                               FALSE                FALSE
## 4                               FALSE                FALSE
## 5                               FALSE                FALSE
## 6                               FALSE                FALSE
##   Feeding.and.Eating.Disorders Sleep.wake.disorders Type.Unspecified
## 1                        FALSE                FALSE            FALSE
## 2                        FALSE                FALSE            FALSE
## 3                        FALSE                FALSE            FALSE
## 4                        FALSE                FALSE            FALSE
## 5                        FALSE                FALSE            FALSE
## 6                        FALSE                FALSE            FALSE
##   Total...of.appts.with.Nurse.Practitioner
## 1                                        0
## 2                                        9
## 3                                        6
## 4                                        5
## 5                                        1
## 6                                        0
##   Total...of.appts.with.Registered.Nurse
## 1                                      0
## 2                                      4
## 3                                      1
## 4                                      2
```

```
## 5                                    14
## 6                                     1
##   Total...of.appts.with.Registered.Dietitian
## 1                                          1
## 2                                          2
## 3                                          1
## 4                                          1
## 5                                          5
## 6                                          1
##   Total...of.appts.with.Social.Worker Total...of.appts.with.Pharmacist
## 1                                   0                                 1
## 2                                   7                                 0
## 3                                   0                                 5
## 4                                   0                                 3
## 5                                   0                                 1
## 6                                   0                                 1
##   Total...of.appts.with.Chiropodist.Wound.Care
## 1                                            0
## 2                                            0
## 3                                            1
## 4                                            2
## 5                                            0
## 6                                            0
##   Total...of.appts.with.Psychologist.Psychiatrist
## 1                                               0
## 2                                               0
## 3                                               0
## 4                                               0
## 5                                               0
## 6                                               0
##   Total...of.appts.with.Kinesiologist Endorefer
## 1                                   0        no
## 2                                   0        no
## 3                                   0        no
## 4                                   0        no
## 5                                   0        no
## 6                                   0        no
```

PART IV - DESRCRIPTIVE DATA ANALYSIS

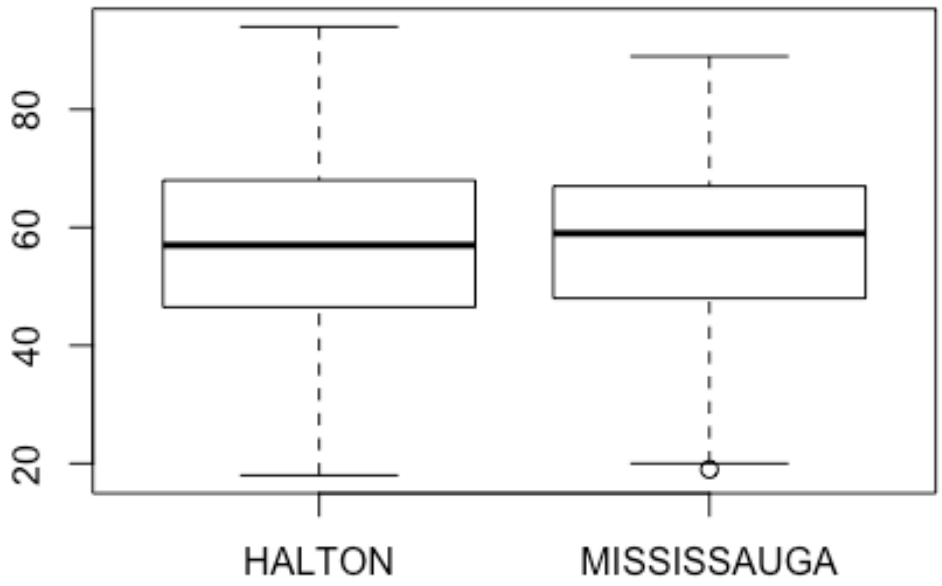
1A.NUMBERS OF MALE AND FEMALE PATIENTS FROM EACH SITE

```
outofregions <- table(dataframe$locationf,dataframe$genderf)
outofregions

##
##                 M   F
##   HALTON       87  69
##   MISSISSAUGA 257 209

prop.table(outofregions,margin=1)

##
##                        M         F
##   HALTON       0.5576923 0.4423077
##   MISSISSAUGA 0.5515021 0.4484979
```

NO EFFECT OF GENDERS - SAME PERCENTAGE OF MALES AND FEMALES FROM BOTH REGIONS

BOXPLOT:AGE VS.LOCATIONS

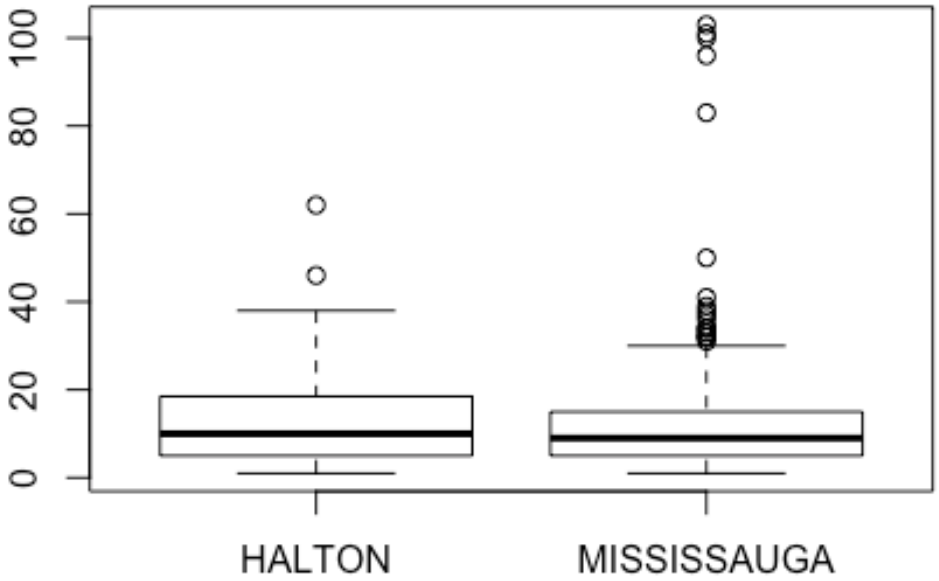
```
boxplot(dataframe$age~dataframe$locationf)
```



BOXPLOT ANALYSIS SUGGEST THAT THERE ARE NO GREAT DIFFERENCE IN MIDIAN AND SPREAD FOR NUMBERS OF PATIENTS WITH DIFFERENT AGES VISITNG BOTH SITES.

BOXPLOT:DISTANCE VS.LOCATION

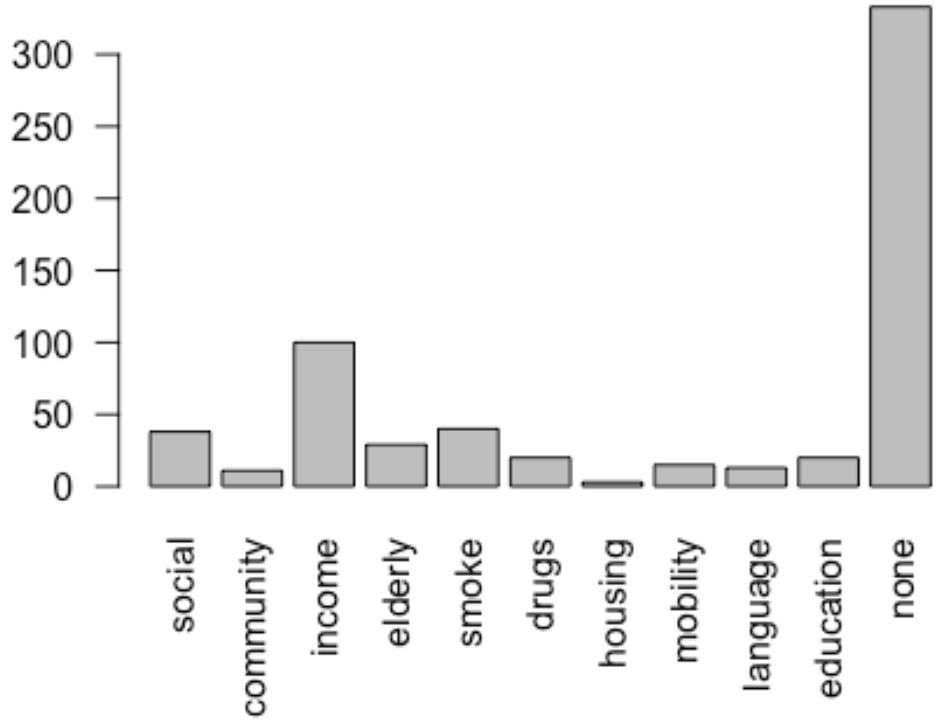
```
boxplot(dataframe$distance~dataframe$locationf)
```



THE MEDIAN DISTANCE TO BOTH SITES ARE VERY SIMILARL; MORE POEPLE WHO NEED TO TRAVEL LONGER DISTANCE TEND TO VISIT MISSISSAUGA SITE.

1B.NUMBERS OF PATIENTS WITH EACH SOCIAL ISSUE/RANGE OF SOCIAL ISSUES:

```
socialissues <- table(dataframe$socialissue)
socialissues

##
##    social community    income   elderly     smoke     drugs   housing
##        38        11       100        29        40        20         3
##  mobility  language education      none
##        15        13        20       333

plot(dataframe$socialissue,las=2)
```



COMPARE TO THE IMPACTFUL ISSUES YOU MENTIONED, THE FOLLOWING SOCIAL ISSUES APPEARS TO BE MORE FREQUENT: LOW INCOME, SOCIAL SUPPORT, SMOKING, ELDERLY

1C.DISTRIBUTION OF CO-MORBIDITY AMONG PATIENTS/NUMBERS OF PATIENTS WITH EACH CO-MORBIDITY

```
avgcomo=subset(dataframe, select=c("Other","Eating.Disorders","Hypertension",
"Cardiovascular.disease","Chronic.kidney.disease","Retinopathy.or.Other.eye.d
isease","X.Non.healing.wounds..greater.than.3.months.",
                            "Neuropathy","Liver.disease..fatty.liver.","Periphe
ral.vascular.disease","Obesity..BMI...30.","Current.malignancy.cancer.treatme
nt","Hyperlipidemia","Thyroid.disease..other.endocrinopathies.POCT..Cushings.
","Dementia",
                                  "Pulmonary.disease..COPD..Asthma.","Obstructive.
sleep.apnea","HIV.AIDS","Pancreas.Diseases","Celiac.Disease","Genetic.Syndrom
es","Deafness.and.or.Blindness","Mental.health"))
```

```
library(tidyr)

## Warning: package 'tidyr' was built under R version 3.1.2

library(dplyr)
avgcomo %>% gather(col,v,Other:Mental.health) %>% table()

##                                                      v
## col                                                  FALSE TRUE
##   Other                                                391  231
##   Eating.Disorders                                     614    8
##   Hypertension                                         228  394
##   Cardiovascular.disease                               481  141
##   Chronic.kidney.disease                               555   67
##   Retinopathy.or.Other.eye.disease                     589   33
##   X.Non.healing.wounds..greater.than.3.months.         610   12
##   Neuropathy                                           600   22
##   Liver.disease..fatty.liver.                          606   16
##   Peripheral.vascular.disease                          584   38
##   Obesity..BMI...30.                                   322  300
##   Current.malignancy.cancer.treatment                  599   23
##   Hyperlipidemia                                       240  382
##   Thyroid.disease..other.endocrinopathies.POCT..Cushings.  551   71
##   Dementia                                             604   18
##   Pulmonary.disease..COPD..Asthma.                     588   34
##   Obstructive.sleep.apnea                              582   40
##   HIV.AIDS                                             622    0
##   Pancreas.Diseases                                    610   12
##   Celiac.Disease                                       620    2
##   Genetic.Syndromes                                    621    1
##   Deafness.and.or.Blindness                            608   14
##   Mental.health                                        436  186
```

THE CO-MORBIDITIES WITH THE HIGHEST FREQUENCY:CARDIOVASCULAR.DISEASE

1C.DISTRIBUTION OF CHRONIC MORBIDITY AMONG PATIENTS

```
sumofcomofoeachpt=apply(avgcomo,1,sum)
summary(sumofcomofoeachpt)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   2.000   3.000   3.288   4.000  10.000
```

EACH PATIENTS HAVE 3 DIFFERENT CO-MORBIDITY ON AVERAGE

1C.NUMBERS OF PATIENTS WITH EACH CHRONIC-MORBIDITY

```
avgchronic=subset(dataframe, select=c("Other.Chronic.","Neuropathy.1","Retino
pathy","Blindness","Cardiovascular.disease.Chronic.","Wounds..non.healing.",
                                "Amputation","Skin.conditions..cutaneous.manifes
tations.","Lipohypertrophy","Hypoglycemia.unawareness",
                                "Diabetic.myonecrosis","Foot.problems..Charcot.s
.","Stiff.man.s.syndrome","Hearing.impairment","Fractures","Nephropathy"))
avgchronic %>% gather(col,v,Other.Chronic.:Nephropathy) %>% table()
```

```
##                                                 v
## col                                              FALSE TRUE
##   Other.Chronic.                                   593   29
##   Neuropathy.1                                     467  155
##   Retinopathy                                      512  110
##   Blindness                                        615    7
##   Cardiovascular.disease.Chronic.                  590   32
##   Wounds..non.healing.                             591   31
##   Amputation                                       615    7
##   Skin.conditions..cutaneous.manifestations.       607   15
##   Lipohypertrophy                                  614    8
##   Hypoglycemia.unawareness                         606   16
##   Diabetic.myonecrosis                             622    0
##   Foot.problems..Charcot.s.                        615    7
##   Stiff.man.s.syndrome                             622    0
##   Hearing.impairment                               621    1
##   Fractures                                        622    0
##   Nephropathy                                      505  117
```

HIGHEST FREQUENCY AMONG CHRONIC-MORBIDITY:NEUROPATHY (155)

1C.AVERAGE # OF CHRONIC-MORBIDITIES

```
sumofchronicfoeachpt=apply(avgchronic,1,sum)
summary(sumofchronicfoeachpt)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.8601  1.0000  6.0000
```

EACH PATIENT HAS AT LEAST 1 CHRONIC-MORBIDITY ON AVERAGE

1C.NUMBERS OF PATIENTS WITH MENTAL HEALTH AS CO-MORBIDITY

```
comomental <- table(dataframe$Mental.health)
comomental

##
## FALSE  TRUE
##   436   186

which(dataframe$Mental)

##   [1]   2   7   9  10  11  26  27  31  32  38  40  41  46  52  56  69  78
##  [18]  80  85  87  94  96  97 104 110 112 122 123 124 127 128 134 135 150
##  [35] 159 161 166 167 169 170 174 176 178 181 190 195 197 198 199 200 204
##  [52] 207 208 210 213 215 227 230 232 241 248 249 251 253 257 259 261 265
##  [69] 266 270 272 273 274 277 279 284 285 291 292 293 295 302 303 316 317
##  [86] 320 324 325 327 329 330 332 340 342 344 345 352 356 357 361 367 368
## [103] 371 374 379 383 387 389 396 399 404 405 409 411 412 416 419 420 421
## [120] 424 425 426 430 432 433 437 446 452 456 457 463 467 468 469 470 475
## [137] 478 479 482 485 487 488 489 490 491 493 498 499 501 503 506 515 517
## [154] 521 524 525 530 531 533 536 538 542 547 550 555 556 558 561 565 568
## [171] 577 582 584 589 590 592 594 595 597 601 605 607 609 614 618 622
```

186 PATIENTS HAVE MENTAL HEALTH AS ONE OF THEIR CO-MORBIDITY, IT ALSO PROVIDES INFORMAITON ABOUT WHO THOSE PATIENTS ARE

1C.HOW MANY PATIENTS HAVE MENTAL HEALTH AS ONE OF THE CO-MORBIDITIES(REGARDING PROTIONS OF F AND M)

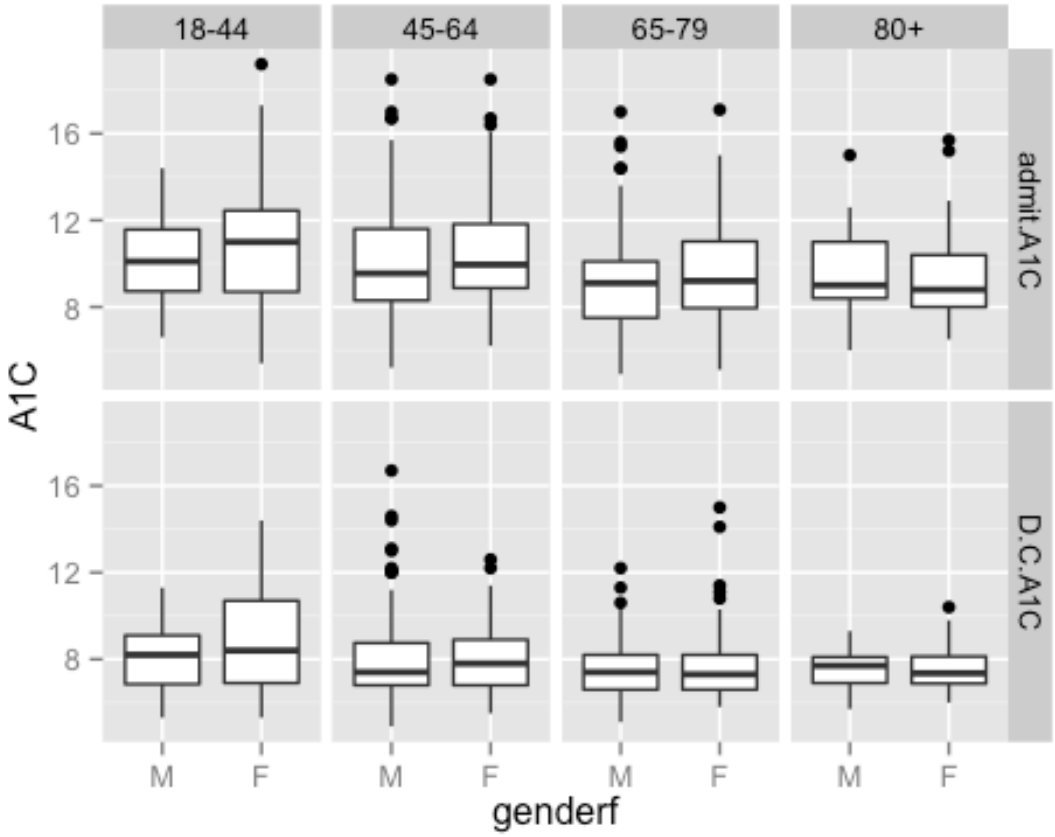
```
comomental <- table(dataframe$Mental.health,dataframe$genderf)
comomental

##
##           M   F
##   FALSE 260 176
##   TRUE   84 102

prop.table(comomental,margin=2)
```

```
##
##                  M         F
##   FALSE 0.7558140 0.6330935
##   TRUE  0.2441860 0.3669065
```

AMONG MALE, ONLY 25% HAVE MENTAL HEALTH ISSUES, BUT HIGHER PROPORTION OF FEMALE THAN MALE HAVE MENTAL HEALTH ISSUE

2A.COMPARING ADMIT AND DISCHARGED CLINICAL METRICS (A1C)

```
library(dplyr)
library(tidyr)
suppressMessages(library(ggplot2))

## Warning: package 'ggplot2' was built under R version 3.1.3

dataframe %>% gather(admitdischarge,A1C,c(admit.A1C,D.C.A1C)) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80+")))->mutatea
  mutatea %>% ggplot(aes(x=genderf,y=A1C))+geom_boxplot()+facet_grid(admitdis
charge~agef)

## Warning in loop_apply(n, do.ply): Removed 1 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 1 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 25 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 45 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 20 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 5 rows containing non-finite
## values (stat_boxplot).
```



DISCHARGED A1C VALUES ARE SUBSTENTICALLY LOWER ACROSS ALL AGE GROUPS THAN ADMIT A1C LEVEL, WHICH SUGGEST THERE IS TREATMENT EFFECT; OLDER PATIENTS TEND TO HAVE LOWER ADMIT AND DISCHARGED A1C VALUE COMPARE TO YOUNGER PATIENTS IN GENERAL; NOT MUCH GENDERS EFFECTS AS THE CHANGES OF A1C VALUES IN MALES AND FEMALES ARE SIMILAR; ALSO, A1C VALUES APPEARS TO DECREASE AS AGE INCREASE FOR BOTH ADIMIT AND DISCHARGE SITUATIONS

SUMMARY OF THE DATA

```
mutatea %>% group_by(admitdischarge,agef,genderf) %>% summarise(Q1=quantile(A
1C,0.25,na.rm=T),
    med=median(A1C,na.rm=T),Q3=quantile(A1C,0.75,na.rm=T),missing=sum(is.na(A
1C)),n=n())

## Source: local data frame [16 x 8]
## Groups: admitdischarge, agef
##
##    admitdischarge  agef genderf    Q1   med     Q3 missing   n
## 1       admit.A1C 18-44       M 8.725 10.10 11.575       0  54
## 2       admit.A1C 18-44       F 8.700 11.00 12.450       0  59
```

```
## 3       admit.A1C 45-64       M 8.300  9.55 11.600       0 172
## 4       admit.A1C 45-64       F 8.875  9.95 11.825       1 125
## 5       admit.A1C 65-79       M 7.500  9.10 10.100       0  97
## 6       admit.A1C 65-79       F 7.950  9.20 11.025       1  73
## 7       admit.A1C   80+       M 8.400  9.00 11.000       0  21
## 8       admit.A1C   80+       F 8.000  8.80 10.400       0  21
## 9         D.C.A1C 18-44       M 6.850  8.20  9.100      11  54
## 10        D.C.A1C 18-44       F 6.900  8.40 10.700      14  59
## 11        D.C.A1C 45-64       M 6.800  7.40  8.750      29 172
## 12        D.C.A1C 45-64       F 6.800  7.80  8.900      16 125
## 13        D.C.A1C 65-79       M 6.600  7.40  8.200       8  97
## 14        D.C.A1C 65-79       F 6.600  7.30  8.200      12  73
## 15        D.C.A1C   80+       M 6.900  7.70  8.100       0  21
## 16        D.C.A1C   80+       F 6.875  7.35  8.125       5  21
```

DATA SHOW THAT BOTH DISCHARGE A1C VALUES ARE LOWER ACROSS AGE GROUP AND GENDER, SUGGESTING SUBSTENTIAL TREATMENT EFFECTS ON DECREASING A1C LEVEL

```
dataframe %>% gather(admitdischarge,A1C,c(admit.A1C,D.C.A1C)) %>%
  filter(is.na(A1C)) %>% select(c(admitdischarge,A1C))
```

```
##    admitdischarge A1C
## 1       admit.A1C  NA
## 2       admit.A1C  NA
## 3         D.C.A1C  NA
## 4         D.C.A1C  NA
## 5         D.C.A1C  NA
## 6         D.C.A1C  NA
## 7         D.C.A1C  NA
## 8         D.C.A1C  NA
## 9         D.C.A1C  NA
## 10        D.C.A1C  NA
## 11        D.C.A1C  NA
## 12        D.C.A1C  NA
## 13        D.C.A1C  NA
## 14        D.C.A1C  NA
## 15        D.C.A1C  NA
## 16        D.C.A1C  NA
## 17        D.C.A1C  NA
## 18        D.C.A1C  NA
## 19        D.C.A1C  NA
## 20        D.C.A1C  NA
## 21        D.C.A1C  NA
## 22        D.C.A1C  NA
```

```
## 23        D.C.A1C  NA
## 24        D.C.A1C  NA
## 25        D.C.A1C  NA
## 26        D.C.A1C  NA
## 27        D.C.A1C  NA
## 28        D.C.A1C  NA
## 29        D.C.A1C  NA
## 30        D.C.A1C  NA
## 31        D.C.A1C  NA
## 32        D.C.A1C  NA
## 33        D.C.A1C  NA
## 34        D.C.A1C  NA
## 35        D.C.A1C  NA
## 36        D.C.A1C  NA
## 37        D.C.A1C  NA
## 38        D.C.A1C  NA
## 39        D.C.A1C  NA
## 40        D.C.A1C  NA
## 41        D.C.A1C  NA
## 42        D.C.A1C  NA
## 43        D.C.A1C  NA
## 44        D.C.A1C  NA
## 45        D.C.A1C  NA
## 46        D.C.A1C  NA
## 47        D.C.A1C  NA
## 48        D.C.A1C  NA
## 49        D.C.A1C  NA
## 50        D.C.A1C  NA
## 51        D.C.A1C  NA
## 52        D.C.A1C  NA
## 53        D.C.A1C  NA
## 54        D.C.A1C  NA
## 55        D.C.A1C  NA
## 56        D.C.A1C  NA
## 57        D.C.A1C  NA
## 58        D.C.A1C  NA
## 59        D.C.A1C  NA
## 60        D.C.A1C  NA
## 61        D.C.A1C  NA
## 62        D.C.A1C  NA
## 63        D.C.A1C  NA
## 64        D.C.A1C  NA
## 65        D.C.A1C  NA
```

```
## 66        D.C.A1C  NA
## 67        D.C.A1C  NA
## 68        D.C.A1C  NA
## 69        D.C.A1C  NA
## 70        D.C.A1C  NA
## 71        D.C.A1C  NA
## 72        D.C.A1C  NA
## 73        D.C.A1C  NA
## 74        D.C.A1C  NA
## 75        D.C.A1C  NA
## 76        D.C.A1C  NA
## 77        D.C.A1C  NA
## 78        D.C.A1C  NA
## 79        D.C.A1C  NA
## 80        D.C.A1C  NA
## 81        D.C.A1C  NA
## 82        D.C.A1C  NA
## 83        D.C.A1C  NA
## 84        D.C.A1C  NA
## 85        D.C.A1C  NA
## 86        D.C.A1C  NA
## 87        D.C.A1C  NA
## 88        D.C.A1C  NA
## 89        D.C.A1C  NA
## 90        D.C.A1C  NA
## 91        D.C.A1C  NA
## 92        D.C.A1C  NA
## 93        D.C.A1C  NA
## 94        D.C.A1C  NA
## 95        D.C.A1C  NA
## 96        D.C.A1C  NA
## 97        D.C.A1C  NA
```

```
c=subset(dataframe, select=c("admit.A1C","D.C.A1C","DischSt"))
c %>% gather(admitdischarge,A1C,c(admit.A1C,D.C.A1C)) %>% filter(is.na(A1C))
```

```
##       DischSt admitdischarge A1C
## 1       death      admit.A1C  NA
## 2  withdrawal      admit.A1C  NA
## 3  withdrawal        D.C.A1C  NA
## 4  withdrawal        D.C.A1C  NA
## 5  withdrawal        D.C.A1C  NA
## 6  withdrawal        D.C.A1C  NA
## 7  withdrawal        D.C.A1C  NA
```

```
## 8  withdrawal        D.C.A1C  NA
## 9  withdrawal        D.C.A1C  NA
## 10 withdrawal        D.C.A1C  NA
## 11      death        D.C.A1C  NA
## 12 withdrawal        D.C.A1C  NA
## 13 withdrawal        D.C.A1C  NA
## 14 withdrawal        D.C.A1C  NA
## 15 withdrawal        D.C.A1C  NA
## 16 withdrawal        D.C.A1C  NA
## 17 withdrawal        D.C.A1C  NA
## 18 withdrawal        D.C.A1C  NA
## 19 withdrawal        D.C.A1C  NA
## 20 withdrawal        D.C.A1C  NA
## 21 withdrawal        D.C.A1C  NA
## 22 withdrawal        D.C.A1C  NA
## 23 withdrawal        D.C.A1C  NA
## 24 withdrawal        D.C.A1C  NA
## 25 withdrawal        D.C.A1C  NA
## 26 withdrawal        D.C.A1C  NA
## 27 withdrawal        D.C.A1C  NA
## 28 withdrawal        D.C.A1C  NA
## 29 withdrawal        D.C.A1C  NA
## 30 withdrawal        D.C.A1C  NA
## 31 withdrawal        D.C.A1C  NA
## 32 withdrawal        D.C.A1C  NA
## 33 withdrawal        D.C.A1C  NA
## 34 withdrawal        D.C.A1C  NA
## 35 withdrawal        D.C.A1C  NA
## 36 withdrawal        D.C.A1C  NA
## 37 withdrawal        D.C.A1C  NA
## 38 withdrawal        D.C.A1C  NA
## 39 withdrawal        D.C.A1C  NA
## 40      death        D.C.A1C  NA
## 41 withdrawal        D.C.A1C  NA
## 42 withdrawal        D.C.A1C  NA
## 43 withdrawal        D.C.A1C  NA
## 44      death        D.C.A1C  NA
## 45 withdrawal        D.C.A1C  NA
## 46      death        D.C.A1C  NA
## 47 withdrawal        D.C.A1C  NA
## 48 withdrawal        D.C.A1C  NA
## 49 withdrawal        D.C.A1C  NA
## 50 withdrawal        D.C.A1C  NA
```

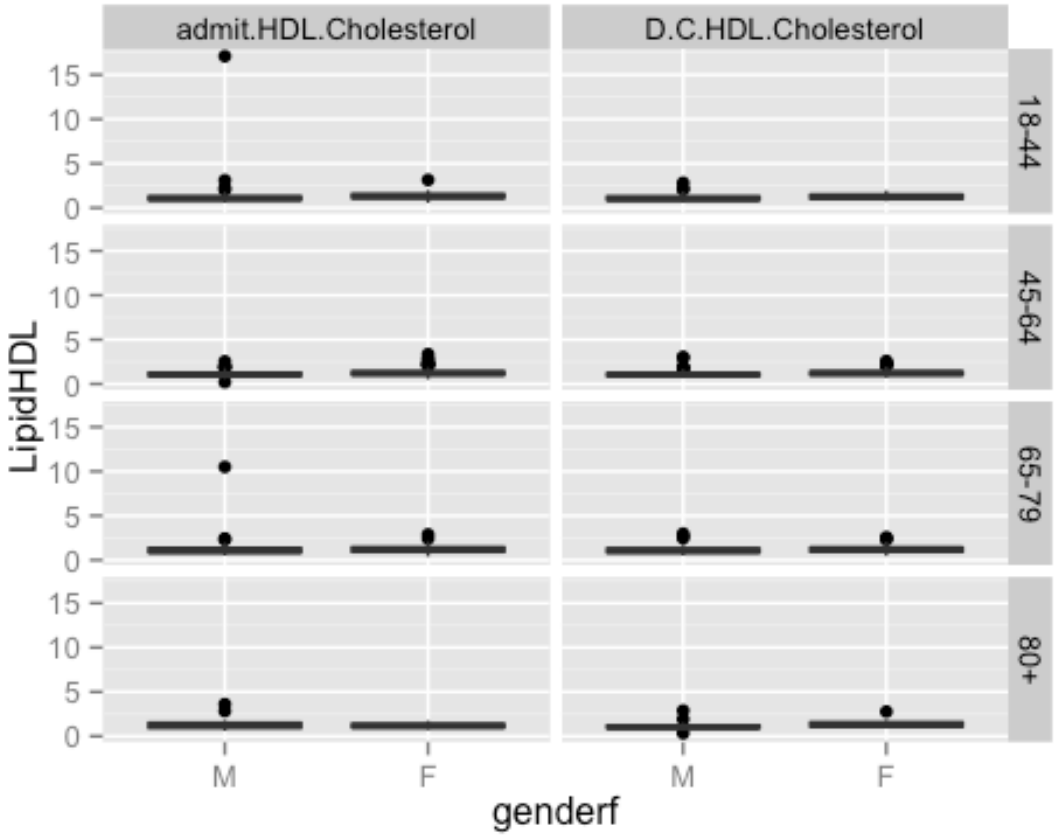
```
## 51 withdrawal        D.C.A1C  NA
## 52 withdrawal        D.C.A1C  NA
## 53 withdrawal        D.C.A1C  NA
## 54 withdrawal        D.C.A1C  NA
## 55 withdrawal        D.C.A1C  NA
## 56 withdrawal        D.C.A1C  NA
## 57      death        D.C.A1C  NA
## 58 withdrawal        D.C.A1C  NA
## 59 withdrawal        D.C.A1C  NA
## 60 withdrawal        D.C.A1C  NA
## 61      death        D.C.A1C  NA
## 62 withdrawal        D.C.A1C  NA
## 63 withdrawal        D.C.A1C  NA
## 64 withdrawal        D.C.A1C  NA
## 65 withdrawal        D.C.A1C  NA
## 66 withdrawal        D.C.A1C  NA
## 67 withdrawal        D.C.A1C  NA
## 68 withdrawal        D.C.A1C  NA
## 69 withdrawal        D.C.A1C  NA
## 70 withdrawal        D.C.A1C  NA
## 71 withdrawal        D.C.A1C  NA
## 72 withdrawal        D.C.A1C  NA
## 73      death        D.C.A1C  NA
## 74 withdrawal        D.C.A1C  NA
## 75 withdrawal        D.C.A1C  NA
## 76 withdrawal        D.C.A1C  NA
## 77 withdrawal        D.C.A1C  NA
## 78 withdrawal        D.C.A1C  NA
## 79 withdrawal        D.C.A1C  NA
## 80 withdrawal        D.C.A1C  NA
## 81 withdrawal        D.C.A1C  NA
## 82 withdrawal        D.C.A1C  NA
## 83 withdrawal        D.C.A1C  NA
## 84 withdrawal        D.C.A1C  NA
## 85 withdrawal        D.C.A1C  NA
## 86      death        D.C.A1C  NA
## 87 withdrawal        D.C.A1C  NA
## 88 withdrawal        D.C.A1C  NA
## 89 withdrawal        D.C.A1C  NA
## 90 withdrawal        D.C.A1C  NA
## 91 withdrawal        D.C.A1C  NA
## 92 withdrawal        D.C.A1C  NA
## 93 withdrawal        D.C.A1C  NA
```

```
## 94 withdrawal        D.C.A1C  NA
## 95 withdrawal        D.C.A1C  NA
## 96 withdrawal        D.C.A1C  NA
## 97 withdrawal        D.C.A1C  NA
```

KEEP IN MIND THAT THERE ARE 97 MISSING A1C VALUES IN TOTAL (2 ARE FROM ADMITED A1C); BUT ALL PATIENTS WHO HAVE MISSING D.C.A1C ARE ACTUALLY WITHDREW BECUASE OF DEATH/WITHDRAWAL, NO MISSING DATA ARE FOUND AMONG THOSE WHO DISCHARGED --> NOTHING TO WORRY ABOUT!

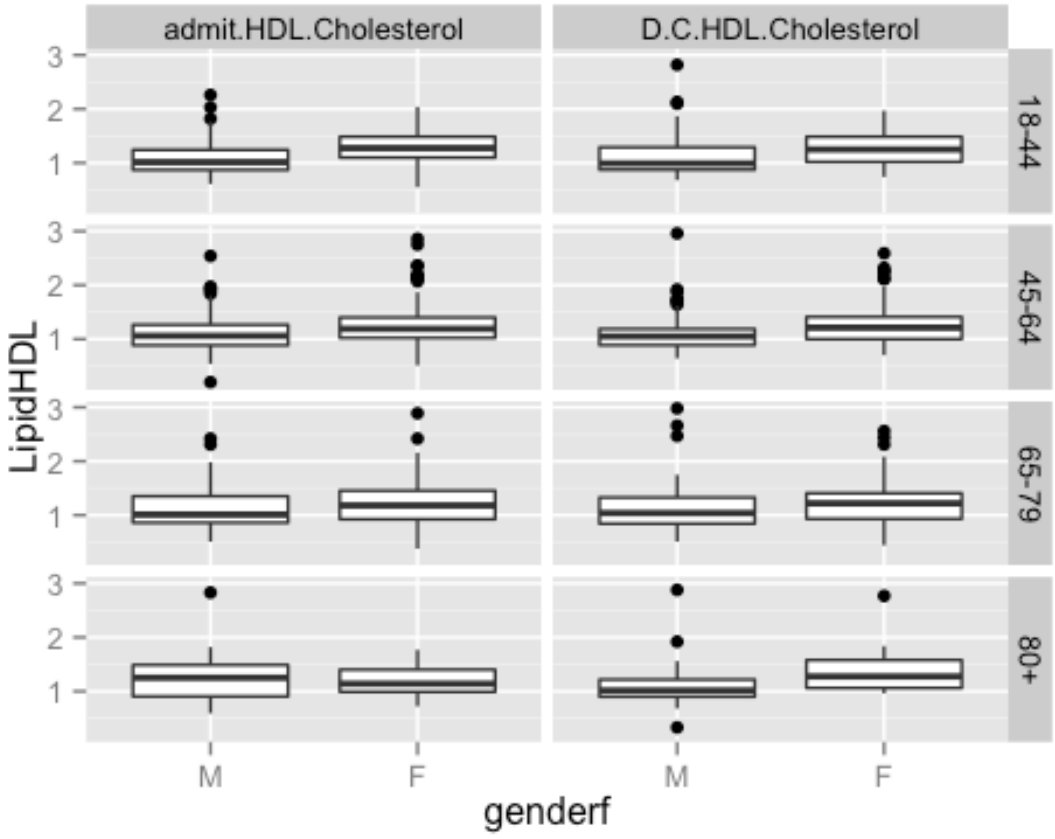
2A.COMPARING ADMIT AND DISCHARGED CLINICAL MATRIX(LIPIDS LEVEL(HDL CHOLESTEROL))

```
dataframe %>% gather(admitdischarge,LipidHDL,c(admit.HDL.Cholesterol, D.C.HDL
.Cholesterol)) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80+")))->mutatea
  mutatea %>% ggplot(aes(x=genderf,y=LipidHDL))+geom_boxplot()+facet_grid(age
f~admitdischarge)

## Warning in loop_apply(n, do.ply): Removed 16 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 36 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 16 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 63 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 12 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 30 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 4 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 8 rows containing non-finite
## values (stat_boxplot).
```



ADDED FILTER TO SELECT HDLLEVEL LESS THAN 3(BECASUE SOME LARGE OUTLIER MADE IT DIFFICULT TO SEE ANY TRENDS ON THE BOXPLOT)

```
mutatea %>% filter(LipidHDL<3) %>%
  ggplot(aes(x=genderf,y=LipidHDL))+geom_boxplot()+facet_grid(agef~admitdisch
arge)
```



SUMMARY OF THE DATA

```
mutatea %>% group_by(admitdischarge,agef,genderf) %>% summarise(Q1=quantile(L
ipidHDL,0.25,na.rm=T),
    med=median(LipidHDL,na.rm=T),Q3=quantile(LipidHDL,0.75,na.rm=T),missing=s
um(is.na(LipidHDL)),n=n())

## Source: local data frame [16 x 8]
## Groups: admitdischarge, agef
##
##           admitdischarge  agef genderf     Q1   med     Q3 missing   n
## 1  admit.HDL.Cholesterol 18-44       M 0.8900 1.030 1.3025       6  54
## 2  admit.HDL.Cholesterol 18-44       F 1.1000 1.280 1.5600      10  59
## 3  admit.HDL.Cholesterol 45-64       M 0.8850 1.060 1.2650       9 172
## 4  admit.HDL.Cholesterol 45-64       F 1.0225 1.190 1.4225       7 125
## 5  admit.HDL.Cholesterol 65-79       M 0.8600 1.010 1.3650       6  97
## 6  admit.HDL.Cholesterol 65-79       F 0.9250 1.180 1.4550       6  73
## 7  admit.HDL.Cholesterol   80+       M 0.9200 1.280 1.5100       0  21
## 8  admit.HDL.Cholesterol   80+       F 0.9900 1.140 1.4000       4  21
## 9    D.C.HDL.Cholesterol 18-44       M 0.8750 0.990 1.2900      16  54
```

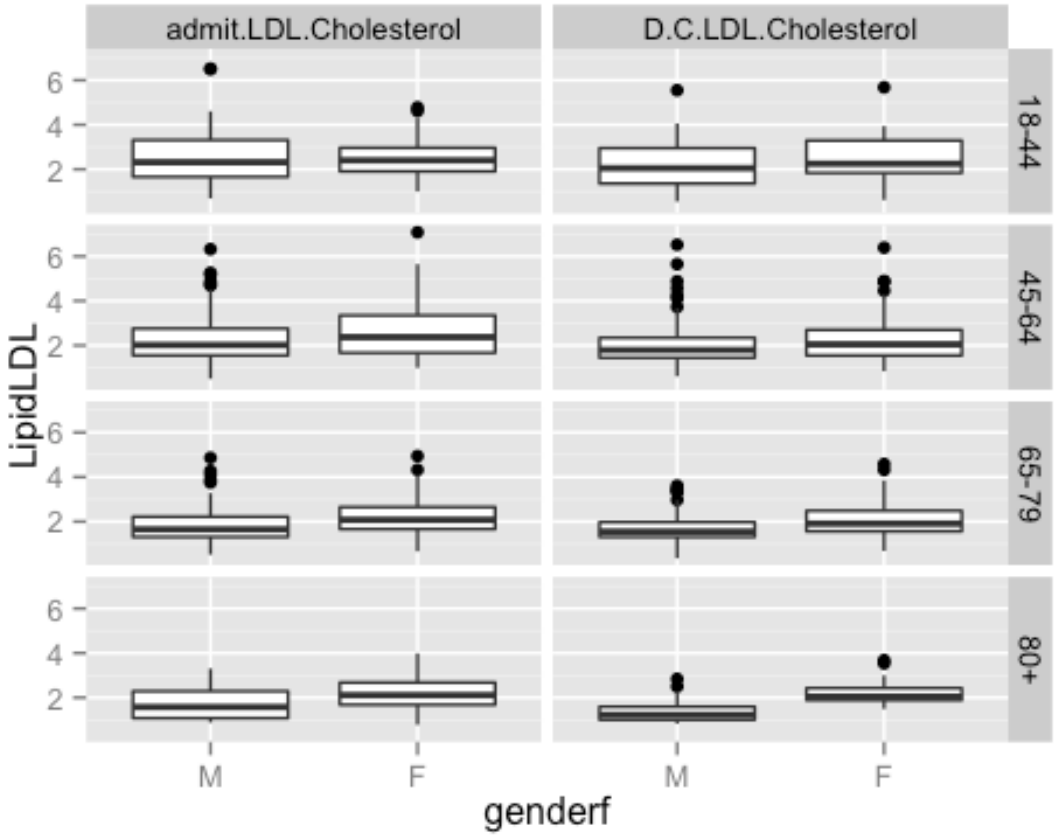
```
## 10   D.C.HDL.Cholesterol 18-44       F 1.0200 1.250 1.4850      20  59
## 11   D.C.HDL.Cholesterol 45-64       M 0.8900 1.050 1.2100      39 172
## 12   D.C.HDL.Cholesterol 45-64       F 1.0000 1.210 1.4100      24 125
## 13   D.C.HDL.Cholesterol 65-79       M 0.8425 1.040 1.3275      11  97
## 14   D.C.HDL.Cholesterol 65-79       F 0.9325 1.220 1.4075      19  73
## 15   D.C.HDL.Cholesterol   80+       M 0.9000 1.015 1.2175       3  21
## 16   D.C.HDL.Cholesterol   80+       F 1.0600 1.275 1.5775       5  21
```

FEMALE HAVE HIGHER HDL LIPID LEVEL IN GENERAL REGARDLESS OF TREATMENT, AGE; DISTRIBUTION OF MALE ARE TYPICALLY BETWEEN 0.9 TO 1.3; AND DISTRUBUTION OF FEMALE IS ABOUT 1 TO 1.4 - THERE IS ONLY GENDER EFFECT AS FEMALE ARE CONSISTENTLY HIGHER, BUT IT'S CONSISTENT BETWEEN ADMIT AND DISCHARGED HDL CHOLESTEROL LEVEL = TREATMENT DOESN'T SEEM TO HAVE EFFECT AS THE CHANGES ARE VERY SMALL.

2A.COMPARING ADMIT AND DISCHARGED CLINICAL MATRIX(LIPIDS LEVEL(LDL CHOLESTEROL))

```
dataframe %>% gather(admitdischarge,LipidLDL,c(admit.LDL.Cholesterol, D.C.LDL
.Cholesterol)) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80+")))->mutatea
  mutatea %>% ggplot(aes(x=genderf,y=LipidLDL))+geom_boxplot()+facet_grid(age
f~admitdischarge)

## Warning in loop_apply(n, do.ply): Removed 21 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 37 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 25 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 71 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 17 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 31 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 3 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 8 rows containing non-finite
## values (stat_boxplot).
```



THE MEDIAN OF BOTH ADMIT AND DISCHARGED LDL CHOLESTEROL LEVEL APPEAR TO BE VERY SIMILAR, SUGGESTING VERY MINOR TREATMENT EFFECT ON THIS CLINICAL MATRIX

SUMMARY OF THE DATA

```
mutatea %>% group_by(admitdischarge,agef,genderf) %>% summarise(Q1=quantile(L
ipidLDL,0.25,na.rm=T),
    med=median(LipidLDL,na.rm=T),Q3=quantile(LipidLDL,0.75,na.rm=T),missing=s
um(is.na(LipidLDL)),n=n())

## Source: local data frame [16 x 8]
## Groups: admitdischarge, agef
##
##          admitdischarge  agef genderf     Q1   med     Q3 missing   n
```

```
## 1  admit.LDL.Cholesterol 18-44       M 1.6600 2.320 3.3125       10  54
## 2  admit.LDL.Cholesterol 18-44       F 1.9075 2.405 2.9650       11  59
## 3  admit.LDL.Cholesterol 45-64       M 1.5550 2.020 2.7550       17 172
## 4  admit.LDL.Cholesterol 45-64       F 1.6700 2.370 3.3500        8 125
## 5  admit.LDL.Cholesterol 65-79       M 1.3075 1.650 2.2150        9  97
## 6  admit.LDL.Cholesterol 65-79       F 1.6600 2.080 2.6500        8  73
## 7  admit.LDL.Cholesterol   80+       M 1.0800 1.580 2.2900        0  21
## 8  admit.LDL.Cholesterol   80+       F 1.6800 2.115 2.6775        3  21
## 9    D.C.LDL.Cholesterol 18-44       M 1.3600 2.050 2.9500       17  54
## 10   D.C.LDL.Cholesterol 18-44       F 1.8400 2.250 3.2950       20  59
## 11   D.C.LDL.Cholesterol 45-64       M 1.4500 1.790 2.3400       43 172
## 12   D.C.LDL.Cholesterol 45-64       F 1.5500 2.050 2.7000       28 125
## 13   D.C.LDL.Cholesterol 65-79       M 1.3100 1.550 1.9700       12  97
## 14   D.C.LDL.Cholesterol 65-79       F 1.5600 1.905 2.4825       19  73
## 15   D.C.LDL.Cholesterol   80+       M 1.0100 1.225 1.5900        3  21
## 16   D.C.LDL.Cholesterol   80+       F 1.8875 2.040 2.4250        5  21
```

THERE IS GENDER EFFECT: FEMALE TEND TO HAVE SLIGHTLY HIGHER LDL CHOLESTEROL LEVE THAN MALE ACROSS ALL AGE GROUP AND AT ADMIT AND DISCHARGE; BUT THE TREATMENT EFFECT APPEARS TO BE VERY MINOR SUGGESTED BY THE VERY SIMILAR ADMIT AND DISCHARGE LDL LEVEL FOR BOTH GENDER AND AGE GROUP

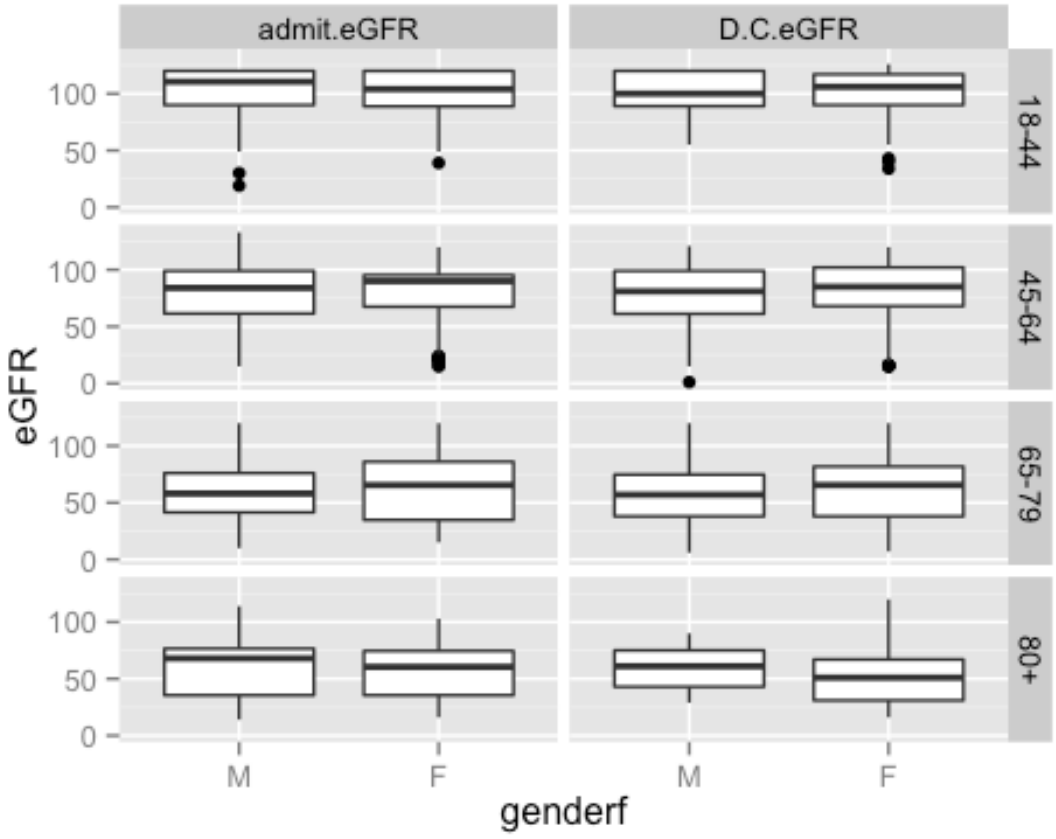
2A.COMPARING ADMIT AND DISCHARGED CLINICAL MATRIX(EGFR)

```
aggregate(cbind(admit.eGFR, D.C.eGFR)~genderf+agef,data=dataframe,mean)
```

```
##   genderf    agef admit.eGFR D.C.eGFR
## 1       M (17,45]  100.34146 96.97561
## 2       F (17,45]   99.55814 98.83721
## 3       M (45,65]   80.70866 78.74055
## 4       F (45,65]   80.74490 79.87755
## 5       M (65,80]   61.97183 58.94366
## 6       F (65,80]   60.26667 58.11111
## 7       M (80,95]   64.15385 56.46154
## 8       F (80,95]   51.50000 49.00000
```

```
dataframe %>% gather(admitdischarge,eGFR,c(admit.eGFR, D.C.eGFR)) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-79","80+")))->mutatea
  mutatea %>% ggplot(aes(x=genderf,y=eGFR))+geom_boxplot()+facet_grid(agef~admitdischarge)
```

```
## Warning in loop_apply(n, do.ply): Removed 15 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 29 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 30 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 66 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 21 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 38 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 5 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 9 rows containing non-finite
## values (stat_boxplot).
```



EGFR SCORES ARE LOWER AS AGE INCREASE; MEDIAN OF DISCHARGED EGFR VALUE DECREASE SUBSTENTICALLY FOR MALE IN AGE GROUP 18-44

SUMMARY OF THE DATA

```
mutatea %>% group_by(admitdischarge,agef,genderf) %>% summarise(Q1=quantile(eGFR,0.25,na.rm=T),
    med=median(eGFR,na.rm=T),Q3=quantile(eGFR,0.75,na.rm=T),missing=sum(is.na(eGFR)),n=n())

## Source: local data frame [16 x 8]
## Groups: admitdischarge, agef
##
##    admitdischarge  agef genderf    Q1   med     Q3 missing   n
## 1      admit.eGFR 18-44       M 90.00 110.5 120.00       8  54
## 2      admit.eGFR 18-44       F 88.75 104.0 120.00       7  59
## 3      admit.eGFR 45-64       M 61.50  84.0  99.00      17 172
## 4      admit.eGFR 45-64       F 67.50  90.0  95.25      13 125
## 5      admit.eGFR 65-79       M 41.50  58.0  76.00      10  97
## 6      admit.eGFR 65-79       F 34.75  65.5  86.00      11  73
```

```
## 7      admit.eGFR   80+       M 35.50  68.0  76.50       2  21
## 8      admit.eGFR   80+       F 35.50  60.5  74.50       3  21
## 9        D.C.eGFR 18-44       M 89.00 100.0 120.00      13  54
## 10       D.C.eGFR 18-44       F 90.00 106.0 117.00      16  59
## 11       D.C.eGFR 45-64       M 61.25  81.0  99.00      38 172
## 12       D.C.eGFR 45-64       F 68.00  85.0 102.00      28 125
## 13       D.C.eGFR 65-79       M 38.00  57.0  74.75      19  97
## 14       D.C.eGFR 65-79       F 38.00  65.5  81.75      19  73
## 15       D.C.eGFR   80+       M 43.00  61.0  75.00       4  21
## 16       D.C.eGFR   80+       F 31.00  51.0  67.00       5  21
```

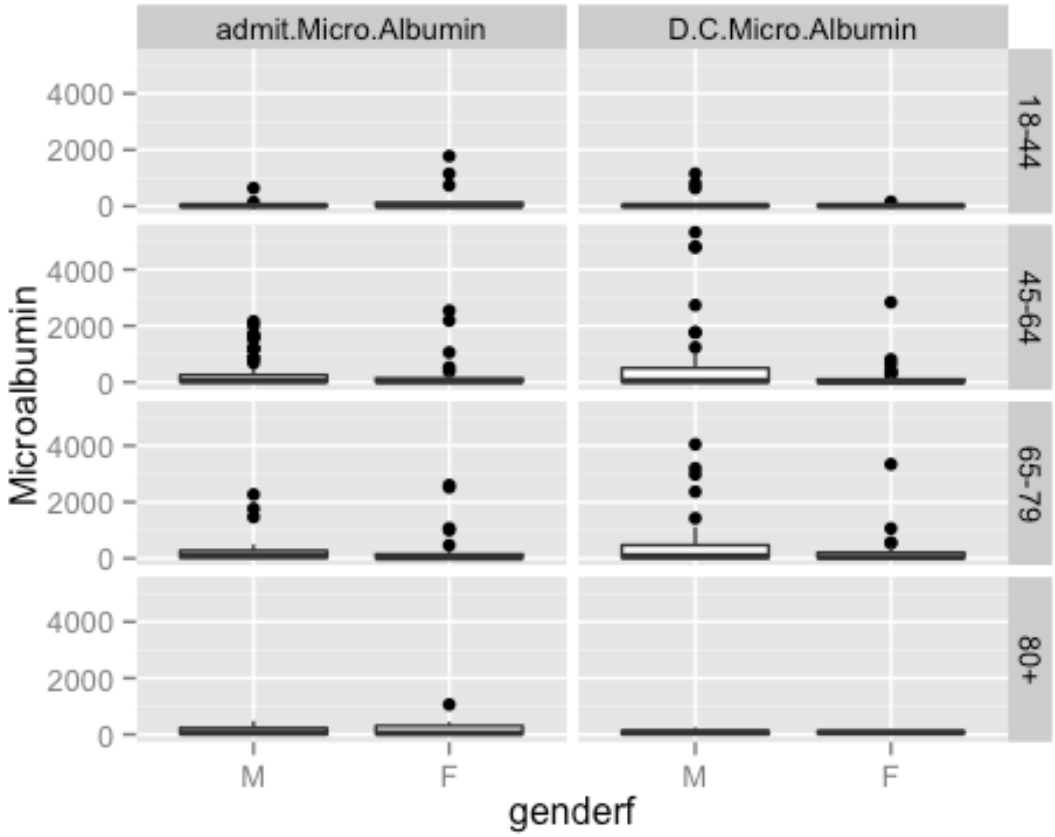
NUMERICAL SUMMARY OF THE DATA: TREATMENT EFFECT APPEARS TO BE SAMLL/NON EXISTENT AS THE DIFFERENCE BETWEEN ADMIT AND DISCHARGE ACROSS GENDER AND AGE IS CONSISTENT; HOWEVER, THERE IS AGE EFFECT: AS AGE INCREASE, EGFR VALUE DECREASE (AT LEAST FOR THE FIRST 3 AGE GROUPS). BUT THERE IS NO CONSISTENT GENDER EFFECT.

2A.COMPARING ADMIT AND DISCHARGED CLINICAL MATRIX(MICRO ALBUMIN)

```
dataframe %>% gather(admitdischarge,Microalbumin,c(admit.Micro.Albumin, D.C.Micro.Albumin)) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-79","80+")))->mutatea
  mutatea %>% ggplot(aes(x=genderf,y=Microalbumin))+geom_boxplot()+facet_grid(agef~admitdischarge)
```

```
## Warning in loop_apply(n, do.ply): Removed 83 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 87 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 183 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 204 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 108 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 108 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 26 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 28 rows containing non-finite
## values (stat_boxplot).
```



#####boxplots in this case are very hard to compare as many of them have outliers. Because of the nature of the Microalbumin level, the range of Microalbumin level can be can be quite large. Thus, need to double check with PI to see which level of filter can be added

SUMMARY OF THE DATA

```
mutatea %>% group_by(admitdischarge,agef,genderf) %>% summarise(Q1=quantile(M
icroalbumin,0.25,na.rm=T),
    med=median(Microalbumin,na.rm=T),Q3=quantile(Microalbumin,0.75,na.rm=T),m
issing=sum(is.na(Microalbumin)),n=n())

## Source: local data frame [16 x 8]
## Groups: admitdischarge, agef
##
##         admitdischarge  agef genderf     Q1    med     Q3 missing   n
## 1  admit.Micro.Albumin 18-44       M  6.700  10.00  35.00      41  54
```

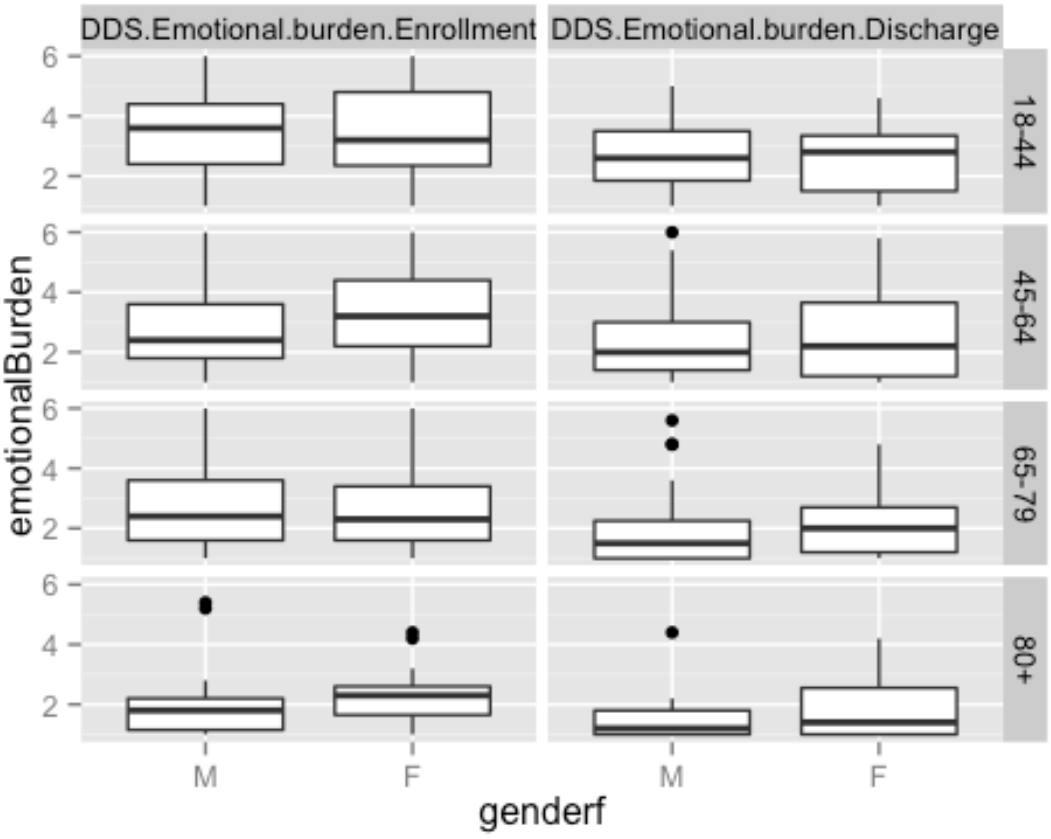
```
## 2  admit.Micro.Albumin 18-44       F  5.000   7.00 121.00      42  59
## 3  admit.Micro.Albumin 45-64       M 12.500  46.00 254.60     105 172
## 4  admit.Micro.Albumin 45-64       F 13.800  32.80 136.00      78 125
## 5  admit.Micro.Albumin 65-79       M 19.400 103.50 274.25      65  97
## 6  admit.Micro.Albumin 65-79       F  8.750  23.95 141.75      43  73
## 7  admit.Micro.Albumin   80+       M  9.400  67.00 228.30      12  21
## 8  admit.Micro.Albumin   80+       F 25.650  46.10 304.50      14  21
## 9    D.C.Micro.Albumin 18-44       M  8.000  13.00  23.00      41  54
## 10   D.C.Micro.Albumin 18-44       F  7.000  11.00  29.08      46  59
## 11   D.C.Micro.Albumin 45-64       M 13.900  47.00 499.00     123 172
## 12   D.C.Micro.Albumin 45-64       F  6.000  16.00  97.75      81 125
## 13   D.C.Micro.Albumin 65-79       M 13.000  77.50 468.05      63  97
## 14   D.C.Micro.Albumin 65-79       F 15.975  39.50 201.10      45  73
## 15   D.C.Micro.Albumin   80+       M 13.650  70.30 122.90      15  21
## 16   D.C.Micro.Albumin   80+       F 45.100  70.00  99.95      13  21
```

THE GENDER AND TREATMENT EFFECTS ARE VERY UNCLEAR BECAUSE THEY ARE VERY INCONSISTENT. THERE MIGHT BE SOME AGE EFFECT. FOR EXAMPLE, YOUNG FEMALE AND MALES PATIENTS TEND TO HAVE LOWER MICROALBUMIN LEVLE FOR DURING ADMIT AND DISCHARGE; OLDER PATIENTS TEND TO HAVE HIGHER MICROALBUMIN LEVEL. HOWEVER, MICROALBUMIN LEVEL STAY ABOUT THE SAME FOR YOUNG PATIENT, WHILE THERE SEEM TO BE A DRAMATIC INCREASE OF DISCHARGE MICROALBUMIN LEVEL FOR BOTH MALE AND FEMALE PATIENTS.

2C.CHANGE IN QOL INDEX (EMOTIONAL BURDEN)

```
dataframe %>% mutate(id=row_number()) %>% gather(admitdischarge,emotionalBurd
en,c(DDS.Emotional.burden.Enrollment,DDS.Emotional.burden.Discharge)) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80+")))->mutatea
  mutatea %>% ggplot(aes(x=genderf,y=emotionalBurden))+geom_boxplot()+facet_g
rid(agef~admitdischarge)

## Warning in loop_apply(n, do.ply): Removed 12 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 75 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 45 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 169 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 12 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 83 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 4 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 20 rows containing non-finite
## values (stat_boxplot).
```



SUMMARY OF THE DATA

```
mutatea %>% group_by(admitdischarge,agef,genderf) %>% summarise(Q1=quantile(e
motionalBurden,0.25,na.rm=T),
    med=median(emotionalBurden,na.rm=T),Q3=quantile(emotionalBurden,0.75,na.r
m=T),missing=sum(is.na(emotionalBurden)),n=n())
```

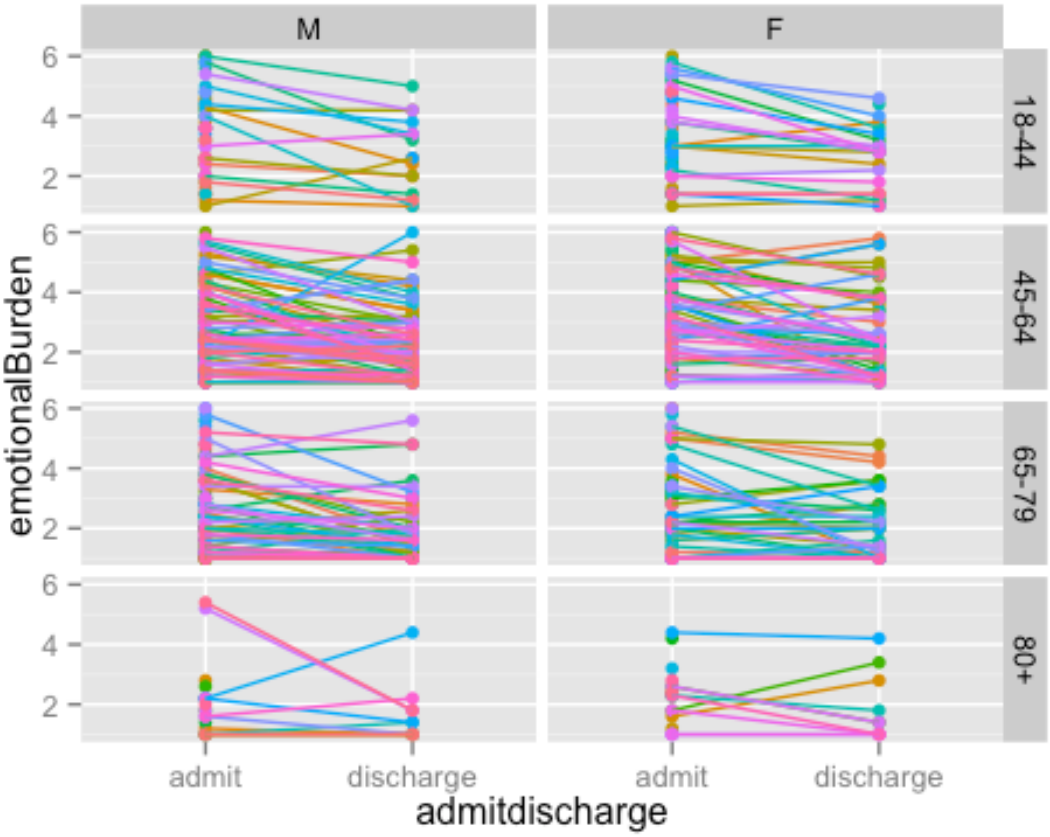
```
## Source: local data frame [16 x 8]
## Groups: admitdischarge, agef
##
##                    admitdischarge  agef genderf   Q1 med   Q3 missing   n
## 1  DDS.Emotional.burden.Enrollment 18-44       M 2.40 3.6 4.40       4  54
## 2  DDS.Emotional.burden.Enrollment 18-44       F 2.35 3.2 4.80       8  59
## 3  DDS.Emotional.burden.Enrollment 45-64       M 1.80 2.4 3.60      29 172
## 4  DDS.Emotional.burden.Enrollment 45-64       F 2.20 3.2 4.40      16 125
## 5  DDS.Emotional.burden.Enrollment 65-79       M 1.60 2.4 3.60       5  97
## 6  DDS.Emotional.burden.Enrollment 65-79       F 1.60 2.3 3.40       7  73
## 7  DDS.Emotional.burden.Enrollment   80+       M 1.15 1.8 2.20       1  21
## 8  DDS.Emotional.burden.Enrollment   80+       F 1.65 2.3 2.60       3  21
## 9   DDS.Emotional.burden.Discharge 18-44       M 1.85 2.6 3.50      38  54
## 10  DDS.Emotional.burden.Discharge 18-44       F 1.50 2.8 3.35      37  59
## 11  DDS.Emotional.burden.Discharge 45-64       M 1.40 2.0 3.00     100 172
## 12  DDS.Emotional.burden.Discharge 45-64       F 1.20 2.2 3.65      69 125
## 13  DDS.Emotional.burden.Discharge 65-79       M 1.00 1.5 2.25      45  97
## 14  DDS.Emotional.burden.Discharge 65-79       F 1.20 2.0 2.70      38  73
## 15  DDS.Emotional.burden.Discharge   80+       M 1.00 1.2 1.80       9  21
## 16  DDS.Emotional.burden.Discharge   80+       F 1.00 1.4 2.55      11  21
```

LOWER EMOTIONAL BURDEN DISCHARGE SCORES FOR BOTH MALE AND FEMALE; FEMALE TEND TO HAVE HIGHER EMOTIONAL BURDEN DISCHARGE SCORES THAN MALE FOR BOTH ADMIT AND DISCHARGED MEASUREMNT (EXCEPT FOR FEMALE IN AGE GROUP 18-44);THE OLDER THE PATIENTS, THE LOWER THE EMOTIONAL BURDEN SCORES ARE MEASURED

SPAGHETTI PLOT

```
  mutatea %>% ggplot(aes(x=admitdischarge,y=emotionalBurden,group=id,color=fa
ctor(id)))+geom_point()+geom_line()+facet_grid(agef~genderf)+guides(color=F)-
>c

  c+scale_x_discrete("admitdischarge",labels=c("DDS.Emotional.burden.Enrollme
nt"="admit","DDS.Emotional.burden.Discharge"="discharge"))

## Warning in loop_apply(n, do.ply): Removed 42 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 45 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 129 rows containing missing
## values (geom_point).
```

```
## Warning in loop_apply(n, do.ply): Removed 85 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 50 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 45 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 10 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 14 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 36 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 33 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 75 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 59 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 42 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 33 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 8 rows containing missing values
## (geom_path).

## Warning in loop_apply(n, do.ply): Removed 8 rows containing missing values
## (geom_path).
```

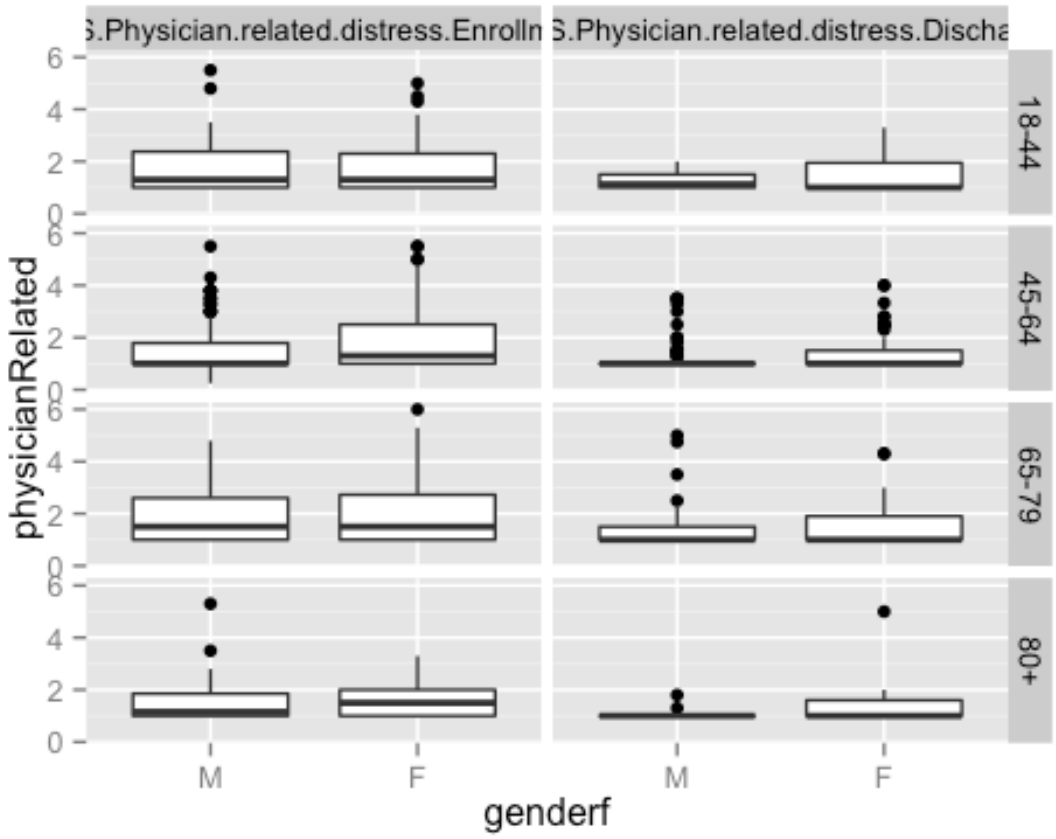


SPAGHETTI PLOT: IT SHOWS THE ADMIT AND DISCHARGE SCRORE FOR EACH INDIVIDUAL; INDICATE TREATMENT EFFECTS AS MOST OF THE LINES GO DOWN (TOWARD BOTTOM RIGHT DIRECTION)

2C.CHANGE IN QOL INDEX (PHYSICIAN RELATED DISTRESS)

```
dataframe %>% mutate(id=row_number()) %>%
  gather(admitdischarge,physicianRelated,c(DDS.Physician.related.distress.Enr
ollment,DDS.Physician.related.distress.Discharge)) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80+")))->mutatea
  mutatea %>% ggplot(aes(x=genderf,y=physicianRelated))+geom_boxplot()+facet_
grid(agef~admitdischarge)

## Warning in loop_apply(n, do.ply): Removed 12 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 75 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 46 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 171 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 13 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 83 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 4 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 20 rows containing non-finite
## values (stat_boxplot).
```



SUMMARY OF THE DATA

```
mutatea %>% group_by(admitdischarge,agef,genderf) %>% summarise(Q1=quantile(p
hysicianRelated,0.25,na.rm=T),
```

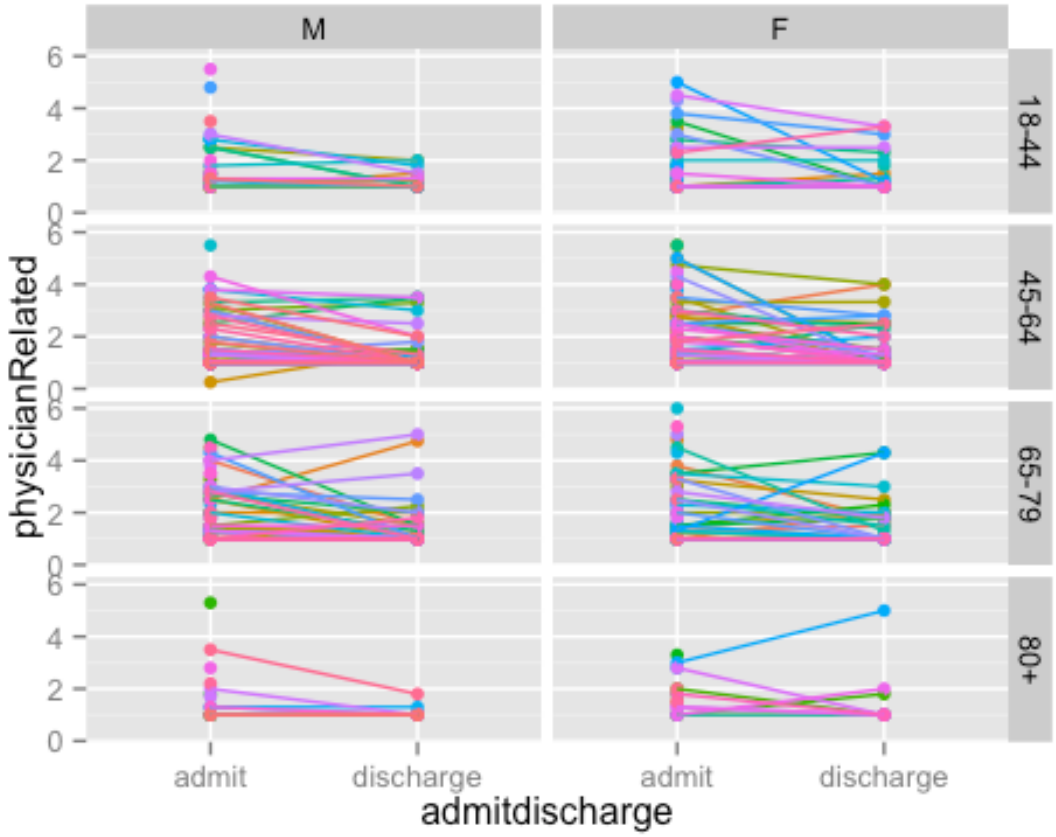
```
    med=median(physicianRelated,na.rm=T),Q3=quantile(physicianRelated,0.75,na
.rm=T),missing=sum(is.na(physicianRelated)),n=n())

## Source: local data frame [16 x 8]
## Groups: admitdischarge, agef
##
##                              admitdischarge  agef genderf Q1   med     Q3
## 1  DDS.Physician.related.distress.Enrollment 18-44       M  1 1.300 2.3750
## 2  DDS.Physician.related.distress.Enrollment 18-44       F  1 1.300 2.3000
## 3  DDS.Physician.related.distress.Enrollment 45-64       M  1 1.000 1.7875
## 4  DDS.Physician.related.distress.Enrollment 45-64       F  1 1.300 2.5000
## 5  DDS.Physician.related.distress.Enrollment 65-79       M  1 1.500 2.6000
## 6  DDS.Physician.related.distress.Enrollment 65-79       F  1 1.500 2.7250
## 7  DDS.Physician.related.distress.Enrollment   80+       M  1 1.150 1.8500
## 8  DDS.Physician.related.distress.Enrollment   80+       F  1 1.500 2.0000
## 9   DDS.Physician.related.distress.Discharge 18-44       M  1 1.125 1.5000
## 10  DDS.Physician.related.distress.Discharge 18-44       F  1 1.000 1.9500
## 11  DDS.Physician.related.distress.Discharge 45-64       M  1 1.000 1.0000
## 12  DDS.Physician.related.distress.Discharge 45-64       F  1 1.000 1.5000
## 13  DDS.Physician.related.distress.Discharge 65-79       M  1 1.000 1.5000
## 14  DDS.Physician.related.distress.Discharge 65-79       F  1 1.000 1.9000
## 15  DDS.Physician.related.distress.Discharge   80+       M  1 1.000 1.0000
## 16  DDS.Physician.related.distress.Discharge   80+       F  1 1.000 1.6000
## Variables not shown: missing (int), n (int)
```

INTERESTINGLY,THE MEDIAN OF THE DISCHARGED PHYSICIAN RELATED DISTRESS LEVEL STAY THE SAME ACROSS GENDER AND AGE GROUP (APPX.1.0); ADMIT PHYSICIAN RELATED DISTRESS LEVEL VARY BY LITTLE BIT, BUT THEY OVERALL HIGHER THAN DISCHARGE DISTRESS LEVEL. THIS SUGGEST TREATMENT EFFECT.

SPAAGETTI PLOT

```
d<-mutatea %>% ggplot(aes(x=admitdischarge,y=physicianRelated,group=id,color=
factor(id))) + geom_point() + geom_line() + facet_grid(agef~genderf)+guides(c
olor=F)

d+scale_x_discrete("admitdischarge",labels=c("DDS.Physician.related.distress.
Enrollment"="admit","DDS.Physician.related.distress.Discharge"="discharge"))

## Warning in loop_apply(n, do.ply): Removed 42 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 45 rows containing missing
## values (geom_point).
```

```
## Warning in loop_apply(n, do.ply): Removed 131 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 86 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 51 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 45 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 10 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 14 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 36 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 33 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 75 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 60 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 41 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 33 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 8 rows containing missing values
## (geom_path).

## Warning in loop_apply(n, do.ply): Removed 8 rows containing missing values
## (geom_path).
```

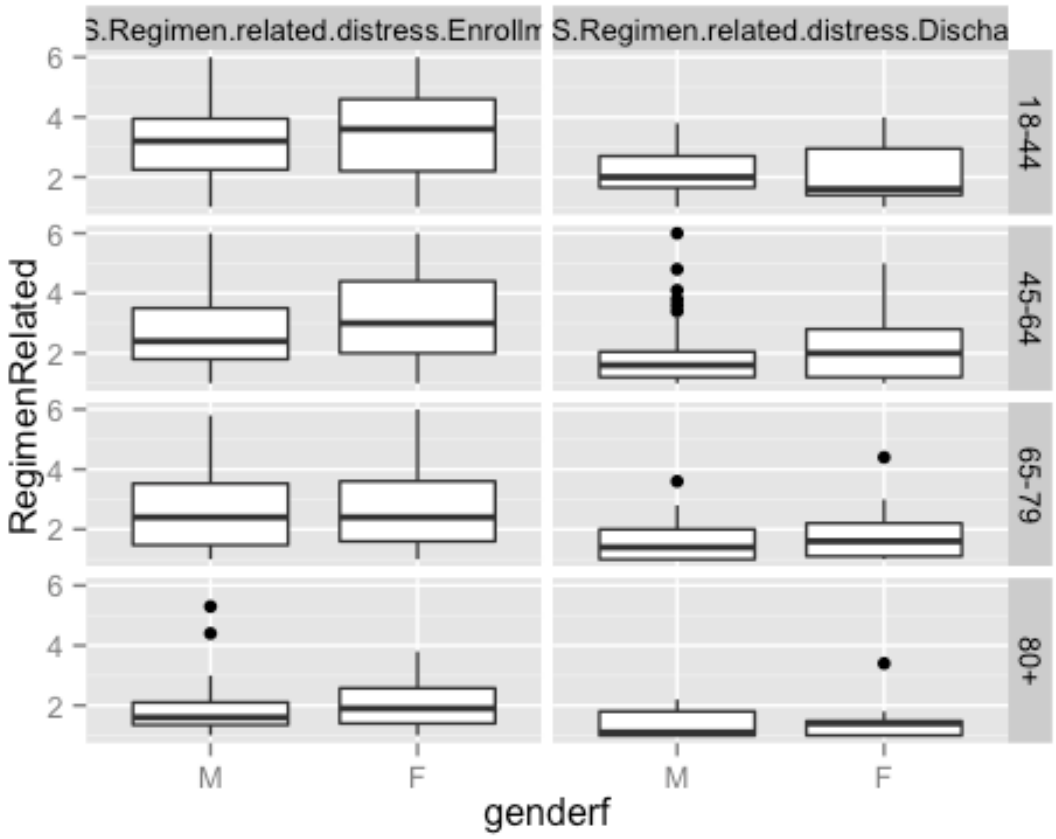


SPAGHETTI PLOT: IT SHOWS THE ADMIT AND DISCHARGE SCRORE FOR EACH INDIVIDUAL; INDICATE TREATMENT EFFECTS AS MOST OF THE LINES GO DOWN (TOWARD BOTTOM RIGHT DIRECTION)

2C.CHANGE IN QOL INDEX (REGIMEN RELATED DISTRESS)

```
dataframe %>% mutate(id=row_number())%>% gather(admitdischarge,RegimenRelated
,c(DDS.Regimen.related.distress.Enrollment,DDS.Regimen.related.distress.Disch
arge)) %>% mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-
64","65-79","80+")))->mutatea
mutatea %>% ggplot(aes(x=genderf,y=RegimenRelated))+geom_boxplot()+facet_grid
(agef~admitdischarge)

## Warning in loop_apply(n, do.ply): Removed 12 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 75 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 45 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 169 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 12 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 83 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 4 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 20 rows containing non-finite
## values (stat_boxplot).
```



SUMMARY OF THE DATA

```
mutatea %>% group_by(admitdischarge,agef,genderf) %>% summarise(Q1=quantile(R
egimenRelated,0.25,na.rm=T),
```

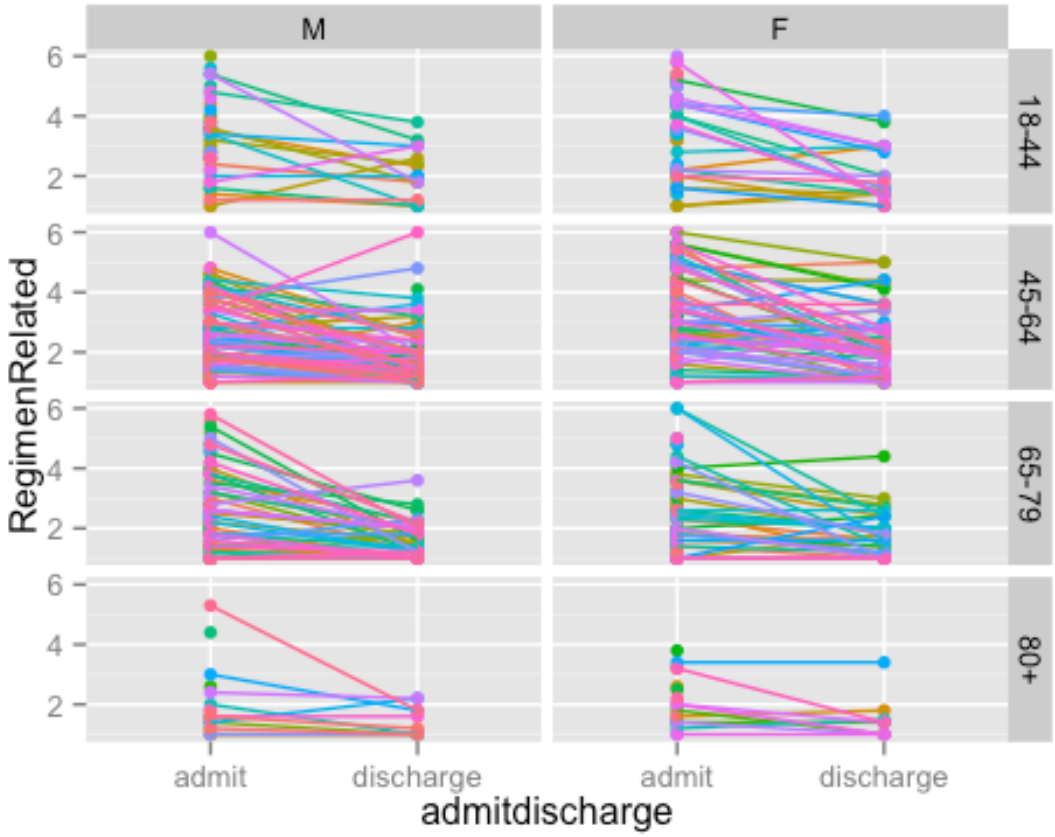
```
    med=median(RegimenRelated,na.rm=T),Q3=quantile(RegimenRelated,0.75,na.rm=
T),missing=sum(is.na(RegimenRelated)),n=n())

## Source: local data frame [16 x 8]
## Groups: admitdischarge, agef
##
##                            admitdischarge  agef genderf    Q1 med    Q3
## 1  DDS.Regimen.related.distress.Enrollment 18-44       M 2.250 3.2 3.950
## 2  DDS.Regimen.related.distress.Enrollment 18-44       F 2.200 3.6 4.600
## 3  DDS.Regimen.related.distress.Enrollment 45-64       M 1.800 2.4 3.500
## 4  DDS.Regimen.related.distress.Enrollment 45-64       F 2.000 3.0 4.400
## 5  DDS.Regimen.related.distress.Enrollment 65-79       M 1.475 2.4 3.525
## 6  DDS.Regimen.related.distress.Enrollment 65-79       F 1.600 2.4 3.600
## 7  DDS.Regimen.related.distress.Enrollment   80+       M 1.350 1.6 2.100
## 8  DDS.Regimen.related.distress.Enrollment   80+       F 1.400 1.9 2.575
## 9   DDS.Regimen.related.distress.Discharge 18-44       M 1.650 2.0 2.700
## 10  DDS.Regimen.related.distress.Discharge 18-44       F 1.400 1.6 2.950
## 11  DDS.Regimen.related.distress.Discharge 45-64       M 1.200 1.6 2.050
## 12  DDS.Regimen.related.distress.Discharge 45-64       F 1.200 2.0 2.800
## 13  DDS.Regimen.related.distress.Discharge 65-79       M 1.000 1.4 2.000
## 14  DDS.Regimen.related.distress.Discharge 65-79       F 1.100 1.6 2.200
## 15  DDS.Regimen.related.distress.Discharge   80+       M 1.000 1.1 1.800
## 16  DDS.Regimen.related.distress.Discharge   80+       F 1.000 1.4 1.475
## Variables not shown: missing (int), n (int)
```

BOTH FEMALE AND MALE PATIENTS ACROSS ALL AGE GROUPS DISPLACE SUBSTENTICALLY LOWER DISCHARGED REGIMEN RELATED DISTRESS SCORE! AMONG THEM, BOTH MALE AND FEMALE PATIENTS IN AGE GROUP 80+ HAVE THE LOWEST ADMIT AND DISCHARGED REGIMEN RELATED DISTRESS SCORE; NO GENDER EFFECTS ARE OBSERVED SINCE THE CHANGES IN ADMIT AND DISCHARGED SCORES ARE SIMILAR FOR BOTH MALE AND FEMALE

SPAAGETTI PLOT

```
  mutatea %>% ggplot(aes(x=admitdischarge,y=RegimenRelated,group=id,color=fac
tor(id)))+geom_point()+geom_line()+facet_grid(agef~genderf)+guides(color=F)->
a

  a+scale_x_discrete("admitdischarge",labels=c("DDS.Regimen.related.distress.
Enrollment"="admit","DDS.Regimen.related.distress.Discharge"="discharge"))

## Warning in loop_apply(n, do.ply): Removed 42 rows containing missing
## values (geom_point).
```

```
## Warning in loop_apply(n, do.ply): Removed 45 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 129 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 85 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 50 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 45 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 10 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 14 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 36 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 33 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 75 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 59 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 42 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 33 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 8 rows containing missing values
## (geom_path).

## Warning in loop_apply(n, do.ply): Removed 8 rows containing missing values
## (geom_path).
```

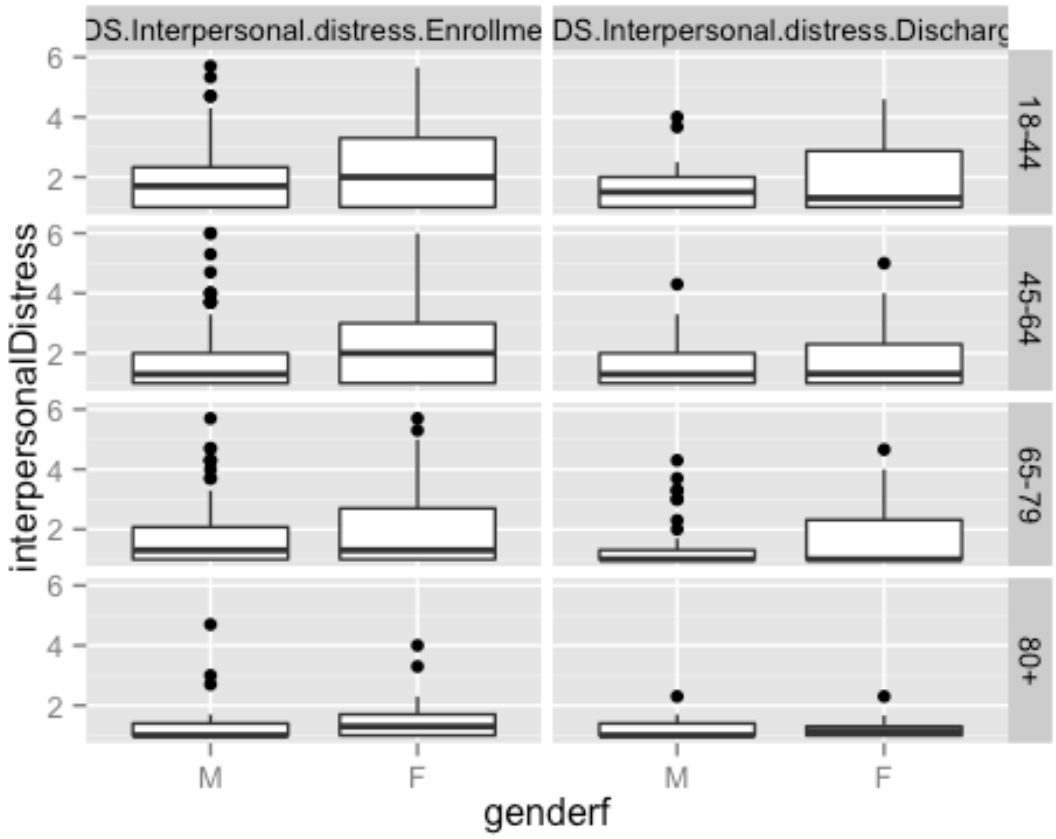


THE SPAAGETTI PLOTS SUGGESTING TREATMENT EFFECTS AS MOST OF THE LINES GO DOWN

2C.CHANGE IN QOL INDEX (INTERPERSONAL DISTRESS)

```
dataframe %>% mutate(id=row_number()) %>% gather(admitdischarge,interpersonal
Distress,c(DDS.Interpersonal.distress.Enrollment,DDS.Interpersonal.distress.D
ischarge)) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80+")))->mutatea
  mutatea %>% ggplot(aes(x=genderf,y=interpersonalDistress))+geom_boxplot()+f
acet_grid(agef~admitdischarge)

## Warning in loop_apply(n, do.ply): Removed 13 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 76 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 47 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 169 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 12 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 83 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 4 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 20 rows containing non-finite
## values (stat_boxplot).
```



SUMMARY OF THE DATA

```
mutatea %>% group_by(admitdischarge,agef,genderf) %>% summarise(Q1=quantile(i
nterpersonalDistress,0.25,na.rm=T),
```

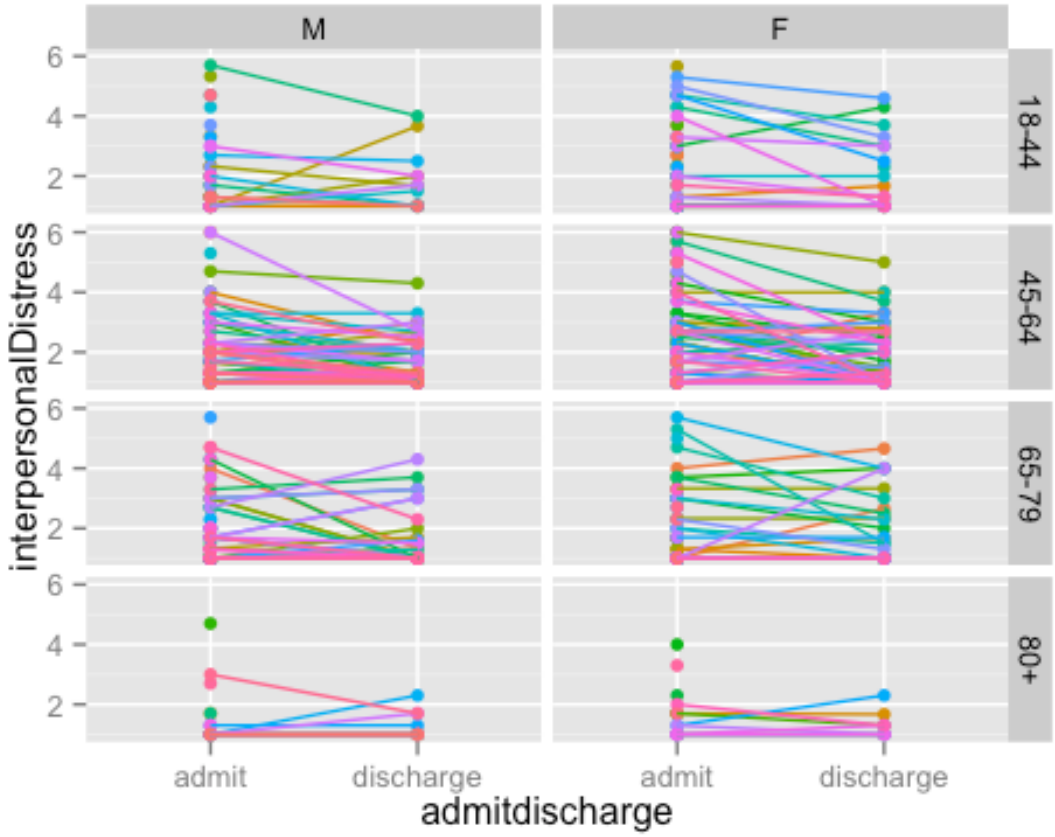
```
    med=median(interpersonalDistress,na.rm=T),Q3=quantile(interpersonalDistre
ss,0.75,na.rm=T),missing=sum(is.na(interpersonalDistress)),n=n())

## Source: local data frame [16 x 8]
## Groups: admitdischarge, agef
##
##                          admitdischarge  agef genderf Q1   med     Q3
## 1  DDS.Interpersonal.distress.Enrollment 18-44       M  1 1.700 2.3300
## 2  DDS.Interpersonal.distress.Enrollment 18-44       F  1 2.000 3.3000
## 3  DDS.Interpersonal.distress.Enrollment 45-64       M  1 1.300 2.0000
## 4  DDS.Interpersonal.distress.Enrollment 45-64       F  1 2.000 3.0000
## 5  DDS.Interpersonal.distress.Enrollment 65-79       M  1 1.300 2.0750
## 6  DDS.Interpersonal.distress.Enrollment 65-79       F  1 1.300 2.7000
## 7  DDS.Interpersonal.distress.Enrollment   80+       M  1 1.000 1.4000
## 8  DDS.Interpersonal.distress.Enrollment   80+       F  1 1.300 1.7000
## 9   DDS.Interpersonal.distress.Discharge 18-44       M  1 1.500 2.0000
## 10  DDS.Interpersonal.distress.Discharge 18-44       F  1 1.300 2.8750
## 11  DDS.Interpersonal.distress.Discharge 45-64       M  1 1.300 2.0000
## 12  DDS.Interpersonal.distress.Discharge 45-64       F  1 1.315 2.3000
## 13  DDS.Interpersonal.distress.Discharge 65-79       M  1 1.000 1.3075
## 14  DDS.Interpersonal.distress.Discharge 65-79       F  1 1.000 2.3150
## 15  DDS.Interpersonal.distress.Discharge   80+       M  1 1.000 1.4000
## 16  DDS.Interpersonal.distress.Discharge   80+       F  1 1.150 1.3000
## Variables not shown: missing (int), n (int)
```

SIMILARLY, BOTH FEMALE AND MALE PATIENTS ACROSS ALL AGE GROUPS DISPLACE SUBSTENTICALLY LOWER INTERPERSONAL DISTRESS SCORE; BOTH MALE AND FEMALE PATIENTS IN AGE GROUP 80+ HAVE THE LOWEST ADMIT AND DISCHARGED INTERPERSONAL DISTRESS SCORE; LEAVE THE OUTLIER

SPAAGETTI PLOT

```
  mutatea %>% ggplot(aes(x=admitdischarge,y=interpersonalDistress,group=id,co
lor=factor(id)))+geom_point()+geom_line()+facet_grid(agef~genderf)+guides(col
or=F)->f

  f+scale_x_discrete("admitdischarge",labels=c("DDS.Interpersonal.distress.En
rollment"="admit","DDS.Interpersonal.distress.Discharge"="discharge"))

## Warning in loop_apply(n, do.ply): Removed 44 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 45 rows containing missing
## values (geom_point).
```

```
## Warning in loop_apply(n, do.ply): Removed 131 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 85 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 50 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 45 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 10 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 14 rows containing missing
## values (geom_point).

## Warning in loop_apply(n, do.ply): Removed 36 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 33 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 75 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 59 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 42 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 33 rows containing missing
## values (geom_path).

## Warning in loop_apply(n, do.ply): Removed 8 rows containing missing values
## (geom_path).

## Warning in loop_apply(n, do.ply): Removed 8 rows containing missing values
## (geom_path).
```



THIS SHOW TREATMENT EFFECT - MOST OF THE LINE GOES DOWN

2D. CHANGE IN CLINICAL METRICS(A1C) FOR SUBGROUPS: MENTAL HEALTH PATIENTS

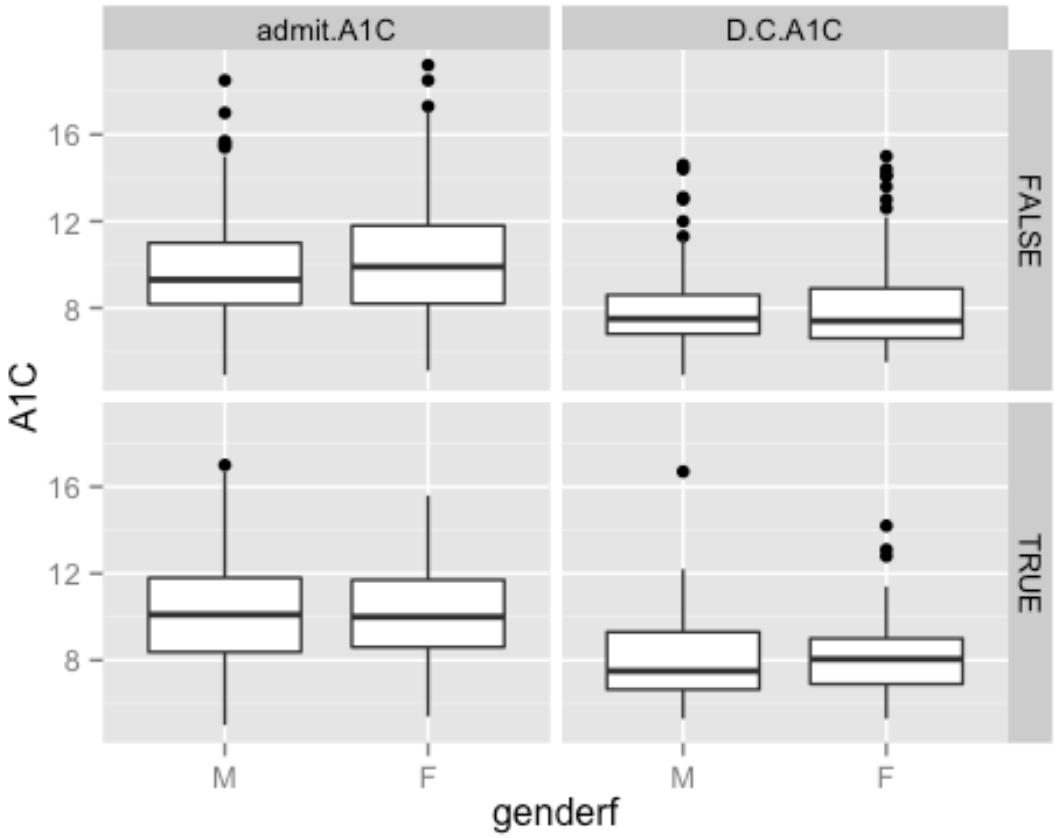
```
dataframe %>% gather(admitdischarge,A1C,c(admit.A1C,D.C.A1C)) %>%
  ggplot(aes(x=genderf,y=A1C))+geom_boxplot()+facet_grid(Mental.health~admitd
ischarge)
```

```
## Warning in loop_apply(n, do.ply): Removed 1 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 58 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 1 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 37 rows containing non-finite
## values (stat_boxplot).
```



DISCHARGED A1C LEVELS ARE SUBSTANTIALLY LOWER THAN ADMIT A1C LEVEL FOR BOTH GROUP OF PATIENTS WITH/WITHOUT MENTAL HEALTH DIAGNOSIS. SUGGESTING THAT THERE IS A TREATMENT EFFCT DUE TO LOWER DISCHAGRED A1C LEVEL,HOWEVER, GENDER AND MENTAL HEALTH DOES NOT HAVE ANY EFFECT ON THE SCORES

SUMMARY OF DATA - WITH DIFFERENCE (DISCHARGE-ADMIT)

```
dataframe %>% mutate(diff=D.C.A1C-admit.A1C) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80")))->mutatea

mutatea %>% group_by(genderf,agef,Mental.health) %>% summarise(Q1=quantile(di
ff,0.25,na.rm=T),
    med=median(diff,na.rm=T),Q3=quantile(diff,0.75,na.rm=T),missing=sum(is.na
(diff)),n=n()) %>% as.data.frame()
```

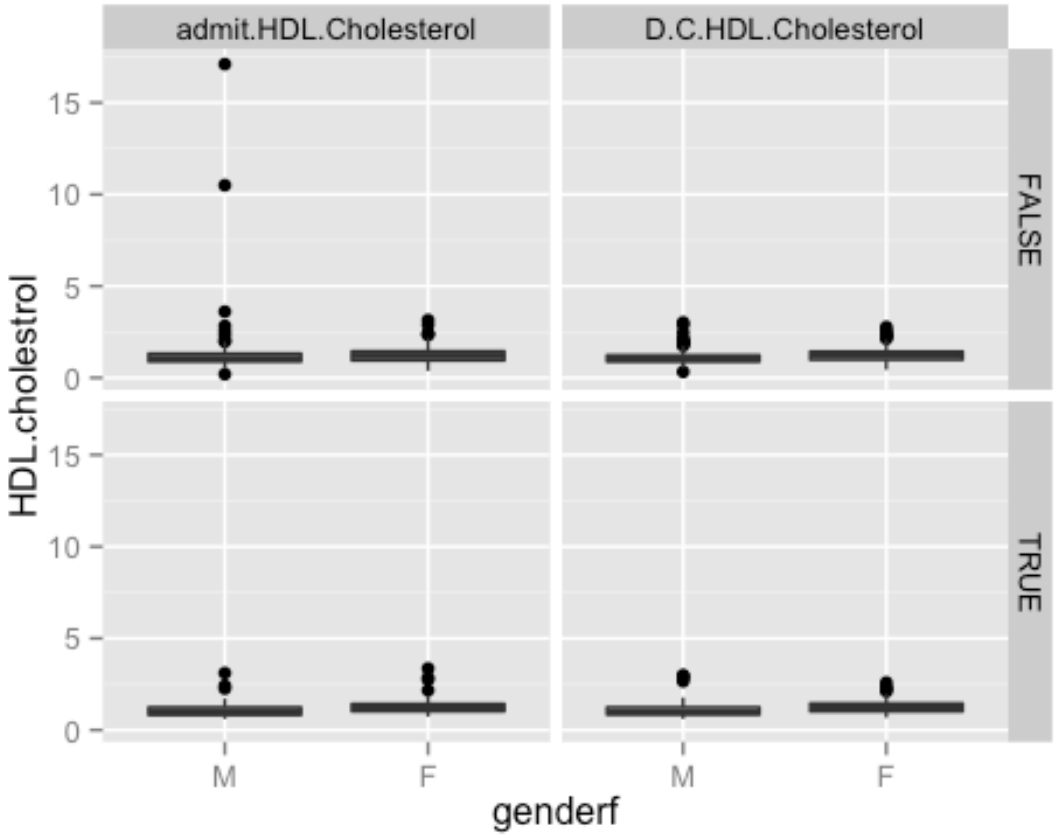
```
##    genderf  agef Mental.health     Q1   med     Q3 missing   n
## 1        M 18-44         FALSE -2.900 -1.50 -0.100       7  36
## 2        M 18-44          TRUE -3.950 -1.90 -0.625       4  18
## 3        M 45-64         FALSE -3.250 -1.90 -0.550      19 122
## 4        M 45-64          TRUE -2.825 -1.90 -0.575      10  50
## 5        M 65-79         FALSE -2.300 -1.20 -0.100       5  82
## 6        M 65-79          TRUE -4.600 -1.20  0.600       3  15
## 7        M    80         FALSE -3.600 -1.40 -0.400       0  20
## 8        M    80          TRUE -2.000 -2.00 -2.000       0   1
## 9        F 18-44         FALSE -4.100 -1.10 -0.025       6  30
## 10       F 18-44          TRUE -2.300 -1.00  0.200       8  29
## 11       F 45-64         FALSE -3.650 -1.85 -0.575       8  72
## 12       F 45-64          TRUE -3.000 -1.90 -0.700       8  53
## 13       F 65-79         FALSE -2.525 -0.90 -0.175      10  58
## 14       F 65-79          TRUE -2.400 -1.70 -0.100       2  15
## 15       F    80         FALSE -5.300 -1.00 -0.600       3  16
## 16       F    80          TRUE -1.200 -0.40 -0.150       2   5
```

NEGATIVE NUMBERS INDICATE TREATMENT EFFECT BECAUSE IT MEANS DISCHARGED A1C LEVEL ARE SMALLER THAN ADMIT A1C LEVELS; HOWEVER, GENDER AND MENTAL HEALTH DOES NOT HAVE ANY EFFECT ON THE SCORES AS THE NEGATIVE VALUES ARE VERY SIMILAR

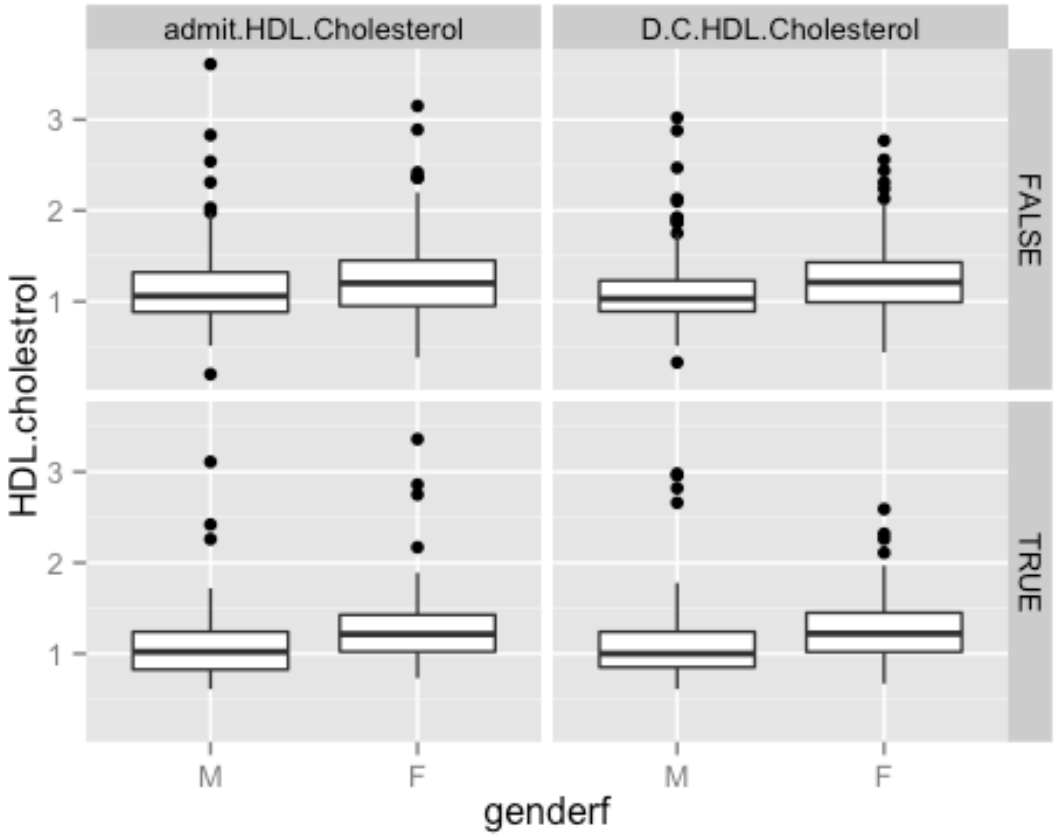
2D. CHANGE IN CLINICAL METRICS(LIPIDS LEVEL(HDL CHOLESTEROL)) FOR SUBGROUPS: MENTAL HEALTH PATIENTS

```
#'with outliers
dataframe %>% gather(admitdischarge,HDL.cholestrol,c(admit.HDL.Cholesterol, D
.C.HDL.Cholesterol)) %>%
  ggplot(aes(x=genderf,y=HDL.cholestrol))+geom_boxplot()+facet_grid(Mental.he
alth~admitdischarge)
```

```
## Warning in loop_apply(n, do.ply): Removed 31 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 93 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 17 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 44 rows containing non-finite
## values (stat_boxplot).
```




```
#'without outliers --> to make it easier to compare the median
dataframe %>% gather(admitdischarge,HDL.cholestrol,c(admit.HDL.Cholesterol, D
.C.HDL.Cholesterol)) %>%
  filter(HDL.cholestrol<5) %>%
  ggplot(aes(x=genderf,y=HDL.cholestrol))+geom_boxplot()+facet_grid(Mental.he
alth~admitdischarge)
```



CHANGES IN HLD.CHOLESTEROL LEVEL SEEMS TO BE QUITE SIMILAR FOR BOTH GROUP OF PATIENTS, WHICH MIGHT SUGGEST THAT TREATMENT MIGHT NOT HAVE MUCH EFFECT ON PATIENT'S HDL.CHOLESTEROL LEVEL

SUMMARY OF DATA - WITH DIFFERENCE (DISCHARGE-ADMIT)

```
dataframe %>% mutate(diff=D.C.HDL.Cholesterol-admit.HDL.Cholesterol) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80")))->mutatea

mutatea %>% group_by(genderf,agef,Mental.health) %>% summarise(Q1=quantile(di
ff,0.25,na.rm=T),
    med=median(diff,na.rm=T),Q3=quantile(diff,0.75,na.rm=T),missing=sum(is.na
(diff)),n=n()) %>% as.data.frame()

##    genderf  agef Mental.health      Q1    med      Q3 missing   n
## 1        M 18-44         FALSE -0.1550 -0.010  0.1050      13  36
## 2        M 18-44          TRUE -0.1225 -0.015  0.0975       6  18
## 3        M 45-64         FALSE -0.1050 -0.010  0.0600      31 122
## 4        M 45-64          TRUE -0.0950  0.035  0.1200      12  50
```

```
## 5        M 65-79         FALSE -0.1750 -0.020  0.0900      11  82
## 6        M 65-79          TRUE -0.1400  0.115  0.2825       3  15
## 7        M    80         FALSE -0.3025 -0.170 -0.0500       2  20
## 8        M    80          TRUE      NA     NA      NA       1   1
## 9        F 18-44         FALSE -0.2050 -0.070  0.0000      12  30
## 10       F 18-44          TRUE -0.2500 -0.150  0.1200      12  29
## 11       F 45-64         FALSE -0.0975  0.000  0.0900      14  72
## 12       F 45-64          TRUE -0.1225 -0.025  0.1125      13  53
## 13       F 65-79         FALSE -0.0900  0.005  0.1350      16  58
## 14       F 65-79          TRUE -0.0700  0.090  0.1900       3  15
## 15       F    80         FALSE -0.0450  0.060  0.2700       5  16
## 16       F    80          TRUE -0.1325 -0.105 -0.0775       3   5
```

ONLY FEW OF THE COMBINATIONS HAVE NEGATIVE VALUES, WHICH MEANS ONLY FEW OF THE CASES WHERE DISCHARGE HDL CHOLESTEROL LEVEL IS SMALLER THAN ADMIT LEVEL. TREATMENT HAS NO EFFECT ON HDL CHOLESTEROL LEVEL FOR BOTH GROUPS OF PATIENTS

2D. CHANGE IN CLINICAL METRICS(LIPIDS LEVEL(LDL CHOLESTEROL)) FOR SUBGROUPS: MENTAL HEALTH PATIENTS

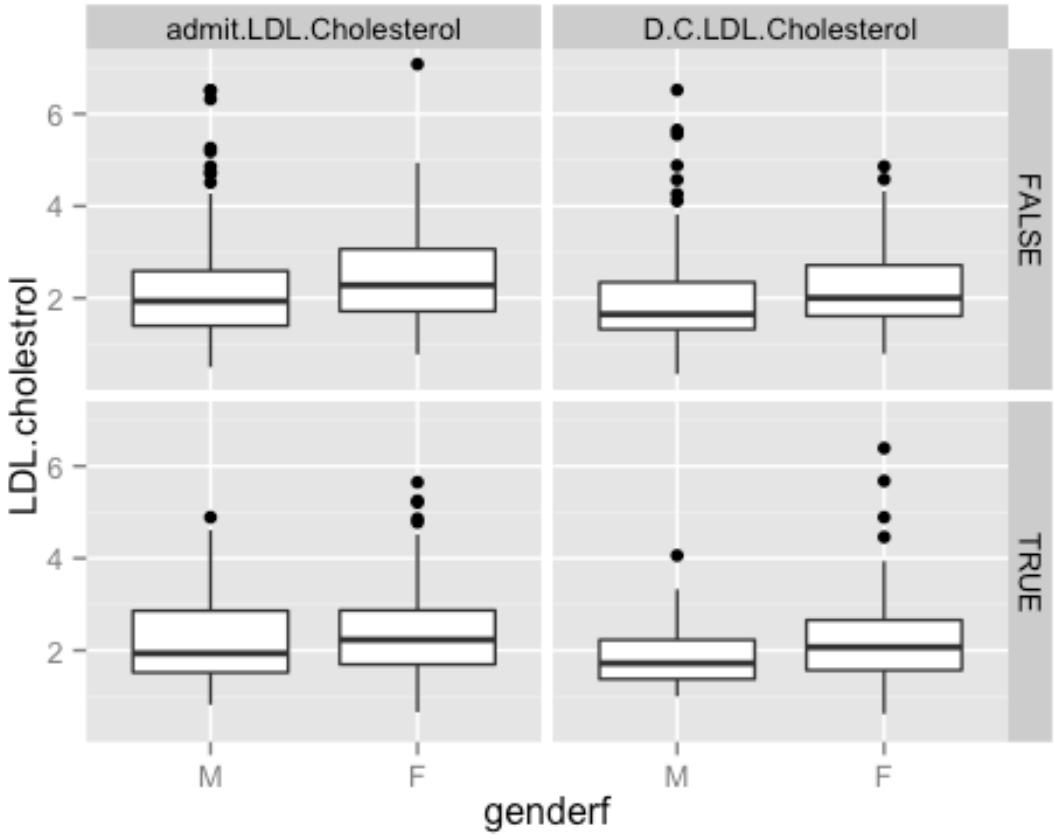
```
dataframe %>% gather(admitdischarge,LDL.cholestrol,c(admit.LDL.Cholesterol, D
.C.LDL.Cholesterol)) %>%
  ggplot(aes(x=genderf,y=LDL.cholestrol))+geom_boxplot()+facet_grid(Mental.he
alth~admitdischarge)
```

```
## Warning in loop_apply(n, do.ply): Removed 40 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 97 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 26 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 50 rows containing non-finite
## values (stat_boxplot).
```



MENTAL HEALTH DOESN'T SEEM TO HAVE AN EFFECT IN THE CHANGES OF LDL CHOLESTEROL LEVEL: THIS IS BECAUSE THE ADMIT LDL CHOLESTEROL FOR BORTH MALE AND FEMALE WITHOUT MENTAL HEALTH DIAGNOSIS DECREASE ABOUT THE SAME RATIO AS THOSE WITH MENTAL HEALTH DIAGNOSIS; HOWEVER, TREATMENT SEEM TO HAVE BIGGER EFFECT FOR MALE PT THAN FEMALE PT AS MALE PT HAVE SUBSTANTIAL LOWER DISCHARGED LDL LIPID LEVEL THAN FEMALE.

SUMMARY OF DATA - WITH DIFFERENCE (DISCHARGE-ADMIT)

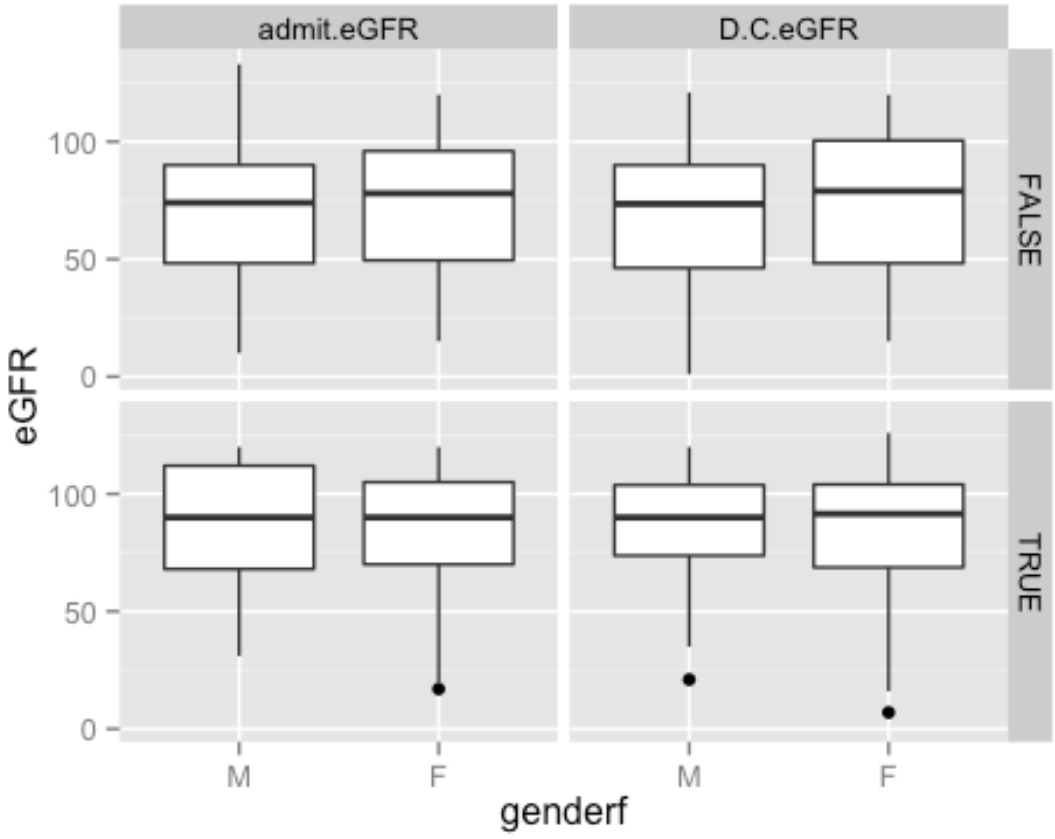
```
dataframe %>% mutate(diff=D.C.LDL.Cholesterol-admit.LDL.Cholesterol) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80")))->mutatea

mutatea %>% group_by(genderf,agef,Mental.health) %>% summarise(Q1=quantile(di
ff,0.25,na.rm=T),
    med=median(diff,na.rm=T),Q3=quantile(diff,0.75,na.rm=T),missing=sum(is.na
(diff)),n=n()) %>% as.data.frame()

##    genderf  agef Mental.health      Q1    med      Q3 missing   n
## 1        M 18-44         FALSE -0.2925 -0.040  0.4675      16  36
```

```
## 2        M 18-44          TRUE -0.8500 -0.395 -0.0325       6  18
## 3        M 45-64         FALSE -0.4900 -0.090  0.2650      35 122
## 4        M 45-64          TRUE -0.7450 -0.350 -0.0600      15  50
## 5        M 65-79         FALSE -0.4475 -0.065  0.2225      14  82
## 6        M 65-79          TRUE -0.1400  0.050  0.1825       3  15
## 7        M    80         FALSE -0.5200 -0.025  0.2025       2  20
## 8        M    80          TRUE      NA     NA      NA       1   1
## 9        F 18-44         FALSE -1.2275  0.010  0.3675      12  30
## 10       F 18-44          TRUE -0.7400 -0.240  0.4300      12  29
## 11       F 45-64         FALSE -0.7200 -0.140  0.2400      15  72
## 12       F 45-64          TRUE -0.7175 -0.135  0.3275      15  53
## 13       F 65-79         FALSE -0.6100 -0.075  0.2850      16  58
## 14       F 65-79          TRUE -0.1650 -0.020  0.1050       4  15
## 15       F    80         FALSE -0.2700  0.005  0.2675       4  16
## 16       F    80          TRUE -0.6075 -0.325 -0.0425       3   5
```

ALTHOUGH MOST OF THE MEDIAN VALUES ARE NEGATIVE (MEANING DISCHARGE IS SMALLER THAN ADMIT), BUT THE DIFFERENCE BETWEEN DISCHARGE AND ADMIT IS VERY SMALL

2D. CHANGE IN CLINICAL METRICS(EGFR) FOR SUBGROUPS: MENTAL HEALTH PATIENTS

```
dataframe %>% gather(admitdischarge,eGFR,c(admit.eGFR, D.C.eGFR)) %>%
  ggplot(aes(x=genderf,y=eGFR))+geom_boxplot()+facet_grid(Mental.health~admit
discharge)

## Warning in loop_apply(n, do.ply): Removed 55 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 96 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 16 rows containing non-finite
## values (stat_boxplot).

## Warning in loop_apply(n, do.ply): Removed 46 rows containing non-finite
## values (stat_boxplot).
```



SUMMARY OF DATA - WITH DIFFERENCE (DISCHARGE-ADMIT)

```
dataframe %>% mutate(diff=D.C.eGFR-admit.eGFR) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80")))->mutatea

mutatea %>% group_by(genderf,agef,Mental.health) %>% summarise(Q1=quantile(di
ff,0.25,na.rm=T),
    med=median(diff,na.rm=T),Q3=quantile(diff,0.75,na.rm=T),missing=sum(is.na
(diff)),n=n()) %>% as.data.frame()

##    genderf  agef Mental.health     Q1   med     Q3 missing   n
## 1        M 18-44         FALSE  -8.00   0.0   5.00      11  36
## 2        M 18-44          TRUE -17.75  -4.5   0.25       6  18
## 3        M 45-64         FALSE  -7.00   0.0   6.00      37 122
## 4        M 45-64          TRUE  -6.75  -2.0   0.75      12  50
## 5        M 65-79         FALSE  -9.00  -2.0   3.00      17  82
## 6        M 65-79          TRUE -13.00  -0.5   4.00       5  15
## 7        M    80         FALSE  -7.00   2.0   4.50       4  20
## 8        M    80          TRUE -22.00 -22.0 -22.00       0   1
```

```
## 9        F 18-44          FALSE -13.50  -2.5   2.00      10  30
## 10       F 18-44           TRUE  -3.00   0.0   7.00       8  29
## 11       F 45-64          FALSE  -7.00   0.0   7.00      17  72
## 12       F 45-64           TRUE  -9.00   0.5   7.00      17  53
## 13       F 65-79          FALSE  -5.00   1.0   4.50      18  58
## 14       F 65-79           TRUE -13.00  -1.0   4.50       4  15
## 15       F    80          FALSE  -5.75  -3.0   6.00       6  16
## 16       F    80           TRUE  -7.00  -6.0  -6.00       2   5
```

MENTAL HEALTH MIGHT BE RELATED TO HIGHER EGFR SCORE: BOTH MALE AND FEMALE WITH MENTAL HEALTH DIAGNOSIS HAVE HIGHER EFGR SCORES THAN THOSE WHO ARE NOT DIAGNOSED WITH MENTAL HEALTH.HOWEVER, THE TREATMENT DOES NOT SEEM TO IMPROVE THE SCORE AS WE SEE THAT DISCHARGED EGFR REMAIN ABOUT THE SAME AS ADMITED EGFR FOR BOTH GENDERS WITH/WITHOUT MENTAL HEALTH DIAGNOSIS.

2D. CHANGE IN CLINICAL METRICS(MICRO ALBUMIN) FOR SUBGROUPS: MENTAL HEALTH PATIENTS

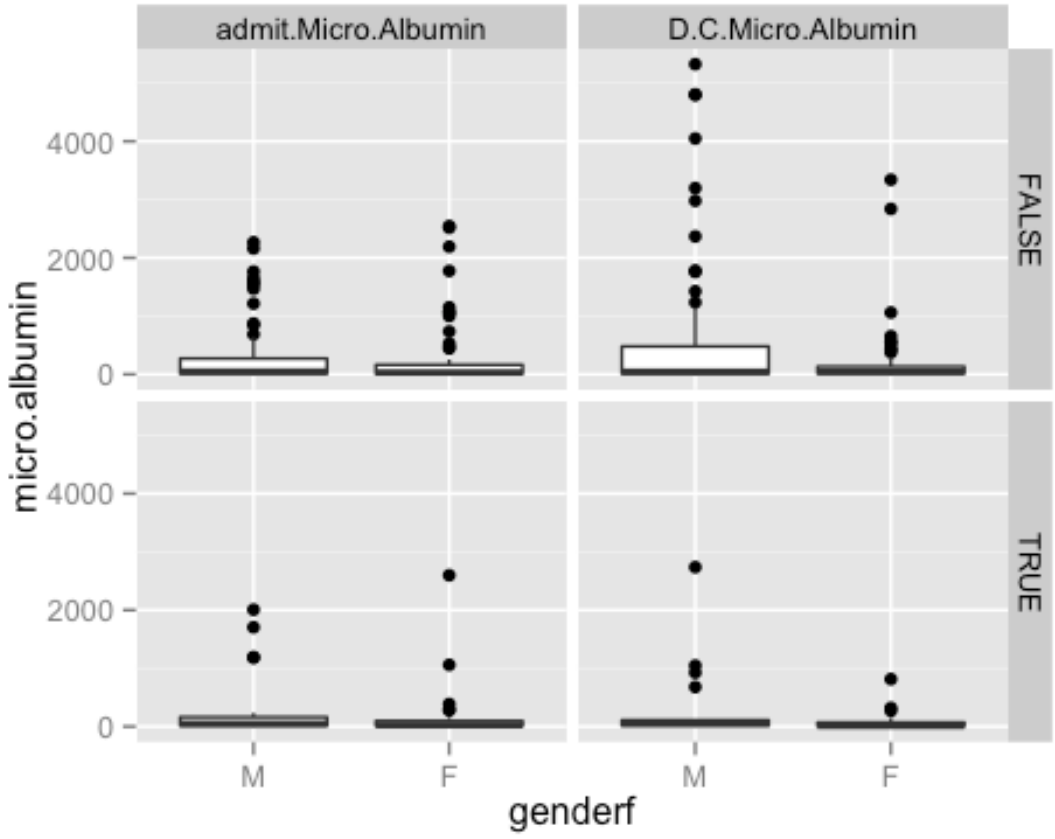
```
dataframe %>% gather(admitdischarge,micro.albumin,c(admit.Micro.Albumin, D.C.
Micro.Albumin)) %>%
  ggplot(aes(x=genderf,y=micro.albumin))+geom_boxplot()+facet_grid(Mental.hea
lth~admitdischarge)
```

```
## Warning in loop_apply(n, do.ply): Removed 276 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 296 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 124 rows containing non-finite
## values (stat_boxplot).
```

```
## Warning in loop_apply(n, do.ply): Removed 131 rows containing non-finite
## values (stat_boxplot).
```



BOXPLOTS IN THIS CASE ARE VERY HARD TO COMPARE AS MANY OF THEM HAVE OUTLIERS. BECAUSE OF THE NATURE OF THE MICROALBUMIN LEVEL, THE RANGE OF MICROALBUMIN LEVEL CAN BE CAN BE QUITE LARGE. THUS, NEED TO DOUBLE CHECK WITH PI TO SEE WHICH LEVEL OF FILTER CAN BE ADDED

SUMMARY OF DATA - WITH DIFFERENCE (DISCHARGE-ADMIT)

```
dataframe %>% mutate(diff=D.C.Micro.Albumin-admit.Micro.Albumin) %>%
  mutate(agef=cut(age,breaks=c(17,44,64,79,95),labels=c("18-44","45-64","65-7
9","80")))->mutatea

mutatea %>% group_by(genderf,agef,Mental.health) %>% summarise(Q1=quantile(di
ff,0.25,na.rm=T),
    med=median(diff,na.rm=T),Q3=quantile(diff,0.75,na.rm=T),missing=sum(is.na
(diff)),n=n()) %>% as.data.frame()

##    genderf  agef Mental.health       Q1    med      Q3 missing   n
## 1        M 18-44         FALSE   -1.925  -0.85   0.000      30  36
## 2        M 18-44          TRUE       NA     NA      NA      18  18
## 3        M 45-64         FALSE  -37.400   2.00 327.500     103 122
```

```
## 4        M 45-64          TRUE  -58.475 -10.45  39.750      40  50
## 5        M 65-79         FALSE  -23.000  12.00 194.925      66  82
## 6        M 65-79          TRUE  -52.750 -27.50  -2.250      13  15
## 7        M    80         FALSE  -59.725 -21.20  15.000      16  20
## 8        M    80          TRUE       NA     NA      NA       1   1
## 9        F 18-44         FALSE  -91.000 -10.70   0.000      25  30
## 10       F 18-44          TRUE    0.450   2.90   5.350      27  29
## 11       F 45-64         FALSE  -50.500  -7.60  27.700      61  72
## 12       F 45-64          TRUE  -37.650  -9.00  -1.700      38  53
## 13       F 65-79         FALSE   -8.325  -0.60   3.250      44  58
## 14       F 65-79          TRUE   -1.500   0.00  11.000      12  15
## 15       F    80         FALSE -249.000 -81.00 -34.250      13  16
## 16       F    80          TRUE   18.000  18.00  18.000       4   5
```

OUTCOMES FOR MALE AND FEMALES ARE DIFFERENT SUGGESTING THAT THERE MIGHT BE A GENDER EFFECT: FOR MALE ALL AGE GROUP WITH MENTAL HEALTH DIAGNOSIS, THEIR DISCHARGE MICROALBUMIN LEVELS ARE SMALLER THAN THEIR ADMIT MICROALBUMIN LEVEL. HOWEVER, PATTERNS ARE NOT CONSISTENT FOR FEMALE PATIENTS. THEREFORE, MENTAL HEALTH DIAGNOSIS MIGHT NOT HAVE ANY IMPACTS IN MICROALBUMIN LEVEL.

PART V - DISCUSSION & CONCLUSION

I am very lucky to have this opportunity to handle such data as an undergrad student. This rare opportunity allows me to utilize my previous knowledge obtained from other statistic classes and apply them here. In particular, both STAC32 and STAC29 classes provided me the necessary Rstudio training and practices required for this project. I wouldn't have the courage and confidence to handle these data without my education from these two classes. In the following session, I am going to discuss in details of what I have learned from both the data and process perspective.

1)WHAT I HAVE LEARNED FROM THE DATA

No surprise, the biggest thing I have learned is data organization (refer to part III for more details of how we handle the obstacles). This experience allows me to have a taste of what's like handling "real-life" data that I normally wouldn't have experienced in a usual classroom setting. That also means lots of efforts have been putting into "cleaning up" the data instead of analyzing data. Although the process could be very tedious and sometimes had me scratch my head, but the feeling of tackling down a big pile of messy data and somehow transforming them into a piece clean and tidy data, is incredible. Along that note, I have also become more familiarize with using many Rstudio build in function that are related to data organizaion. For example, "gather" and "%" (pipe) help us to organize data more efficiently as these two functions play an essential role in every graph we plotted.

I have also gained more experience in writing functions. Before, I always had trouble with "thinking in a computer language", but with many practices from this project, I found myself more comfortable with writing a function or at least thinking in a programming logic. For example, I have become more familiar with defining what data variables are needed inside the function to accomplish its goal or deciding on the set of steps that the function will use to accomplish a particular goal.

Furthermore, I am amazed by how powerful ggplot could be. Apparently, the NewYork times think so too! I have gained so much respect for ggplot from this project mostly because it makes it very easy for us to obtain different custom plots compare to basic graphs. For example, multivariate exploration is greatly simplified through faceting and coloring, which is especially useful when comparing admit and discharged between certain variables (ie., A1C level) with age groups and gender at the same time. By using "facet_grid()", we were able to create a grid of plots consist of all those variables in one place. Also, making such complicated plot doesn't require us to change any of the original plotting code, we simply just need to add the facet command to it. When it comes to continuous variables or factor variables with many levels, coloring by that variable does the trick. Again, ggplot makes this very easy! (ie., the spaghetti plot we made to show the admit and discharge score for each patient distinguished by color).

2)WHAT I HAVE LEARNED FROM THE PROCESS

This is literally the best university/learning experience I have ever had during my time at UTSC. Not only I had the rare opportunity to work on a research project that is going to be published in the near future, but also I have gained so much more experience with Rstudio. Besides, I have learned many things other than data throughout this process. First, doing good data science requires high degrees of paying attention to details, which is the most important thing that I have learned. With 622 subjects and over 50 different variables, that means many potential mistakes could be made. Whether it is reading the wrong data frame, or entering the incorrect variables, attention to details is very important in preventing errors and improving efficiency. Funny story: one time we

spent a good two hours on trying to fix a function because we thought that was where the poblem exist. It turned out it's because I read in the wrong excel file!

Another thing I learned is communication. Good communication skill is the key solution to many problems. Because of the "real life" data we are handling here, there are a lot of times we encountered issues such as missing data, incorrect data, or problems that can not simply be solved by programming. I had to learn to communicate the problem with the principle investigator in a simple and clear way.

I have also learned to challenge myself by stepping out of my "comfort zone". Before, whenever I got stuck in problems that I can not solve, I will just give up or directly ask for an answer to that question. Now I have learned to challenge myself by taking an extra step to think through the problem on my own first; or I have learned to "experiment" with something simple to see if it works out, then apply it to the problem. Alought this process might take up a lot of time and efforts, and it could be "uncomfortable" at first, but it is a necessary step to take because this mindset has taught me to become a better problem solver.

Overall, I truly believe that the technical skills and other skills that I developed through this project would definitely have tremendous benefits in many aspects.

PART VI - RECOMMENDATIONS

Finally, I have listed several analysis ideas that I hope to explore in the future.

1)Analysis of variance - test to see if any differences/change we observe in boxplot are significant or not.

2)Evaluate the cost effectiveness of this program by looking at the changes in average number of ER and hospital admits per month.

3)Evaluate program utilization by looking at: a.numbers of referrals and source of referrals. b.How many patients were enrolled, discharged, died or withdrew from the program during data period. c. The average length of stay, the average length of time between referral to the first visit. d. numbers of urgent or semi-urgent, the capacity of the program (manual extraction, not on MOH data). 3.use PACIC scores to determine how satisfy are the patient with the program.

4)Determine clinician utilization by looking at which clinician do patients see most frequently.