

IST 687 Project Group 1

Catherine Devine, Yodit Ayalew, Lauren Foltz, Jack Harris

```
#Install Packages
#install.packages("flexdashboard")
#install.packages("gdata")
#install.packages("moments")
#install.packages("ggplot2")
#install.packages("ggmap")
#install.packages("zipcode")
#install.packages("maps")
#install.packages("mapproj")
#install.packages("ggcorrplot")

#Install Libraries
library(flexdashboard)
library(gdata)

library(moments)
library(ggplot2)
library(ggmap)

library(zipcode)
library(maps)
library(mapproj)
library(ggcorrplot)

#Read Data
DAPCancerEvents <-
data.frame(read.xls("data/DAP_cancer_events_hosp.xls"),stringsAsFactors=FALSE)
#head(DAPCancerEvents)
#str(DAPCancerEvents)
#summary(DAPCancerEvents)

#Clean data
##Remove first row (NA data)
DAPCancerEvents <- DAPCancerEvents[-1,]
##Remove last row (aggregate United States data)
DAPCancerEvents <- DAPCancerEvents[-939,]
##Remove columns 7 and 8 (confidence interval data)
DAPCancerEvents <- DAPCancerEvents[, -7:-8]
##Remove columns 8 and 9 (confidence interval data)
DAPCancerEvents <- DAPCancerEvents[, -8:-9]
##Remove columns 9 and 10 (confidence interval data)
DAPCancerEvents <- DAPCancerEvents[, -9:-10]
##Remove columns 10 and 11 (confidence interval data)
DAPCancerEvents <- DAPCancerEvents[, -10:-11]
##Remove columns 11 and 12 (confidence interval data)
DAPCancerEvents <- DAPCancerEvents[, -11:-12]
##Remove columns 12 and 13 (confidence interval data)
```

```

DAPCancerEvents <- DAPCancerEvents[, -12:-13]
##Remove columns 13 and 14 (confidence interval data)
DAPCancerEvents <- DAPCancerEvents[, -13:-14]
##Remove columns 14 and 15 (confidence interval data)
DAPCancerEvents <- DAPCancerEvents[, -14:-15]
##Remove columns 15 and 16 (confidence interval data)
DAPCancerEvents <- DAPCancerEvents[, -15:-16]
##Remove columns 18 and 19 (confidence interval data)
DAPCancerEvents <- DAPCancerEvents[, -16:-17]
##Remove columns 19 and 20 (confidence interval data)
DAPCancerEvents <- DAPCancerEvents[, -17:-18]
namelist <- c("Provider ID", "Hospital", "City", "State",
              "Deaths", "% died in hospital",
              "% admitted last month", "Hospital days last month",
              "% admitted ICU last month", "ICU days last month",
              "% life sust treatment last month", "% chemo last 2 weeks",
              "% hospice last month", "hospice days last month",
              "% hospice last 3 days", "% > 10 physicians last 6 months")
names(DAPCancerEvents) <- namelist
#View(DAPCancerEvents)
#str(DAPCancerEvents)

#Define Numberize function to convert to numerics and remove spaces and commas
Numberize<-function(inputVector){
  inputVector<-gsub(",", "", inputVector)
  inputVector<-gsub(" ", "", inputVector)
  return(as.numeric(inputVector)) }

#Convert columns 5 (Deaths) through 16 (% >10 physicians last 6 months) to numbers using
Numberize, as well as remove spaces
DAPCancerEvents$Deaths <- Numberize(DAPCancerEvents$Deaths)
DAPCancerEvents$`% died in hospital` <- Numberize(DAPCancerEvents$`% died in hospital`)
DAPCancerEvents$`% admitted last month` <- Numberize(DAPCancerEvents$`% admitted last
month`)
DAPCancerEvents$`Hospital days last month` <- Numberize(DAPCancerEvents$`Hospital days
last month`)
DAPCancerEvents$`% admitted ICU last month` <- Numberize(DAPCancerEvents$`% admitted ICU
last month`)
DAPCancerEvents$`ICU days last month` <- Numberize(DAPCancerEvents$`ICU days last month`)
DAPCancerEvents$`% life sust treatment last month` <- Numberize(DAPCancerEvents$`% life
sust treatment last month`)
DAPCancerEvents$`% chemo last 2 weeks` <- Numberize(DAPCancerEvents$`% chemo last 2
weeks`)
DAPCancerEvents$`% hospice last month` <- Numberize(DAPCancerEvents$`% hospice last
month`)
DAPCancerEvents$`hospice days last month` <- Numberize(DAPCancerEvents$`hospice days last
month`)
DAPCancerEvents$`% hospice last 3 days` <- Numberize(DAPCancerEvents$`% hospice last 3
days`)
DAPCancerEvents$`% > 10 physicians last 6 months` <- Numberize(DAPCancerEvents$`% > 10
physicians last 6 months`)
#str(DAPCancerEvents)

```

```

#subset Deaths for descriptive function and remove NAs
DAPCancerEvents5 <- DAPCancerEvents[,5]
DAPCancerEvents5 <- na.omit(DAPCancerEvents5)
#subset % died in hospital for descriptive function and remove NAs
DAPCancerEvents6 <- DAPCancerEvents[,6]
DAPCancerEvents6 <- na.omit(DAPCancerEvents6)
#subset % admitted last month for descriptive function and remove NAs
DAPCancerEvents7 <- DAPCancerEvents[,7]
DAPCancerEvents7 <- na.omit(DAPCancerEvents7)
#subset % hospital days last month for descriptive function and remove NAs
DAPCancerEvents8 <- DAPCancerEvents[,8]
DAPCancerEvents8 <- na.omit(DAPCancerEvents8)
#subset % admitted last month for descriptive function and remove NAs
DAPCancerEvents9 <- DAPCancerEvents[,9]
DAPCancerEvents9 <- na.omit(DAPCancerEvents9)
#subset % ICU days last month for descriptive function and remove NAs
DAPCancerEvents10 <- DAPCancerEvents[,10]
DAPCancerEvents10 <- na.omit(DAPCancerEvents10)
#subset % life sustaining treatment last month for descriptive function and remove NAs
DAPCancerEvents11 <- DAPCancerEvents[,11]
DAPCancerEvents11 <- na.omit(DAPCancerEvents11)
#subset % chemo last 2 weeks for descriptive function and remove NAs
DAPCancerEvents12 <- DAPCancerEvents[,12]
DAPCancerEvents12 <- na.omit(DAPCancerEvents12)
#subset % hospice last month for descriptive function and remove NAs
DAPCancerEvents13 <- DAPCancerEvents[,13]
DAPCancerEvents13 <- na.omit(DAPCancerEvents13)
#subset hospice days last month for descriptive function and remove NAs
DAPCancerEvents14 <- DAPCancerEvents[,14]
DAPCancerEvents14 <- na.omit(DAPCancerEvents14)
#subset % hospice last 3 days for descriptive function and remove NAs
DAPCancerEvents15 <- DAPCancerEvents[,15]
DAPCancerEvents15 <- na.omit(DAPCancerEvents15)
#subset % >10 physicians last 6 months for descriptive function and remove NAs
DAPCancerEvents16 <- DAPCancerEvents[,16]
DAPCancerEvents16 <- na.omit(DAPCancerEvents16)

#Color Palettes (sourced from http://www.cookbook-r.com/Graphs/Colors_(ggplot2)/)
#Palette of colors for use in graphs
cbbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2",
"#D55E00", "#CC79A7")

#functions
#create functions to generate descriptive statistics, maps and graphs
descriptive <- function(x){
  print(paste("Mean:",mean(x)))
  print(paste("Median:",median(x)))
  print(paste("Standard Deviation:",sd(x)))
  print(paste("Skewness:",skewness(x)))
  print(paste("Kurtosis:",kurtosis(x)))
  hist(x,main=paste(title),xlab=title,col = cbbPalette[4])
}

xlabel <- "State"
xlabel1 <- "Hospital"

```

```

byhospital <- function(x){
  hospitalmap <-
ggplot(DAPCancerEvents,aes(x=DAPCancerEvents$Hospital,y=col,colour=col)) +
geom_point(colour=cbbPalette[8]) + ggtitle(title1)
  hospitalmap <- hospitalmap + labs(x=xlabel1,y=title)
  hospitalmap <- hospitalmap + theme_classic()
  hospitalmap
}

bystate <- function(x){
  statemap <- ggplot(DAPCancerEvents,aes(x=DAPCancerEvents$State,y=col,
                                         colour=col)) + geom_point(colour=cbbPalette[6]) +
                                         ggtitle(title2)
  statemap <- statemap +labs(x=xlabel,y=title)
  statemap <- statemap + theme_classic()
  statemap
}

#create function to generate maps
mapvariable <- function(x){
  us <- map_data("state")
  map <- ggplot(DAPCancerEvents,aes(map_id=statename))
  map <- map + geom_map(map=us,aes(fill=col))
  map <- map + labs(x=NULL, y=NULL)
  map <- map + expand_limits(x=
    us$long, y=us$lat)
  map <- map + coord_map()
  map <- map + ggtitle(title2)
  map <- map + theme_classic()
  map
}

#state data
#Load zipcode data
data(zipcode)
zipcode1 <- tolower(zipcode$state)
#add statename field
DAPCancerEvents$state <- tolower(DAPCancerEvents$State)
DAPCancerEvents <- DAPCancerEvents[, -4]
#DAPCancerEvents

#create dataframe with state abbreviations and names
statescd <- as.data.frame(state.abb,state.name)
#take rownames and create new field statename from them
statescd$statename <- rownames(statescd)
#set both columns to lowercase
statescd$state.abb <- tolower(statescd$state.abb)
statescd$state <- statescd$state.abb
statescd$statename <- tolower(statescd$statename)
#create dataframe that merges statescd and zip.df by state
DAPCancerEvents <- merge(DAPCancerEvents,statescd, by="state")
#DAPCancerEvents

```

```

#Linear Regression for all models with correlation greater than |0.5|
#regression model for predicting % died in hospital by percentage hospital days last
month
#as the variable
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`% died in
hospital`~DAPCancerEvents$`Hospital days last month`,
                        data=DAPCancerEvents)
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)
#summary(DAPCancerEvents.lm)

#paste("adjusted r squared:",DAPCancerEvents.model$adj.r.squared) #adj r squared

```

Summary

This reflects an analysis of the DAP Cancer Events dataset. This dataset contains 938 rows (one row per hospital), and 12 attributes for each hospital.

We've conducted the following analysis for each attribute

- defined descriptive statistics
- histogram
- scatterplot of data by hospital
- scatterplot of data by state
- map data

In addition, we've run a correlation analysis and then a linear regression model for all correlations greater than |0.5|

Analysis

Column

Deaths

Chart 1

```

#generate descriptive statistics and histograms for each variable
col <- DAPCancerEvents$Deaths
title <- "Deaths"
title1 <- "Deaths by Hospital"
title2 <- "Deaths by State"
#Call descriptive function
descriptive(DAPCancerEvents5)

## [1] "Mean: 145.265458422175"
## [1] "Median: 125"
## [1] "Standard Deviation: 69.6487827302734"
## [1] "Skewness: 2.66519410583897"
## [1] "Kurtosis: 14.3827129459792"

```

Deaths

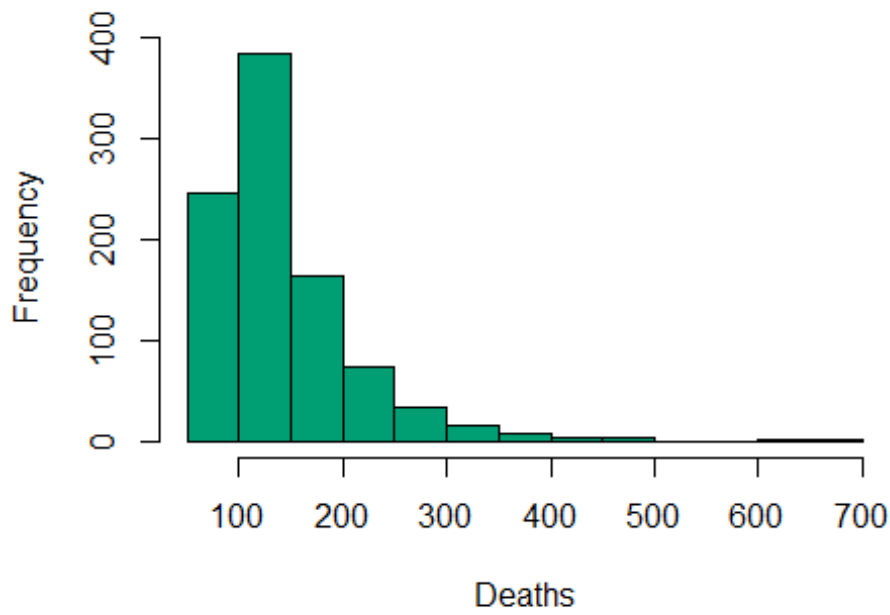


Chart 2

`byhospital(col)`

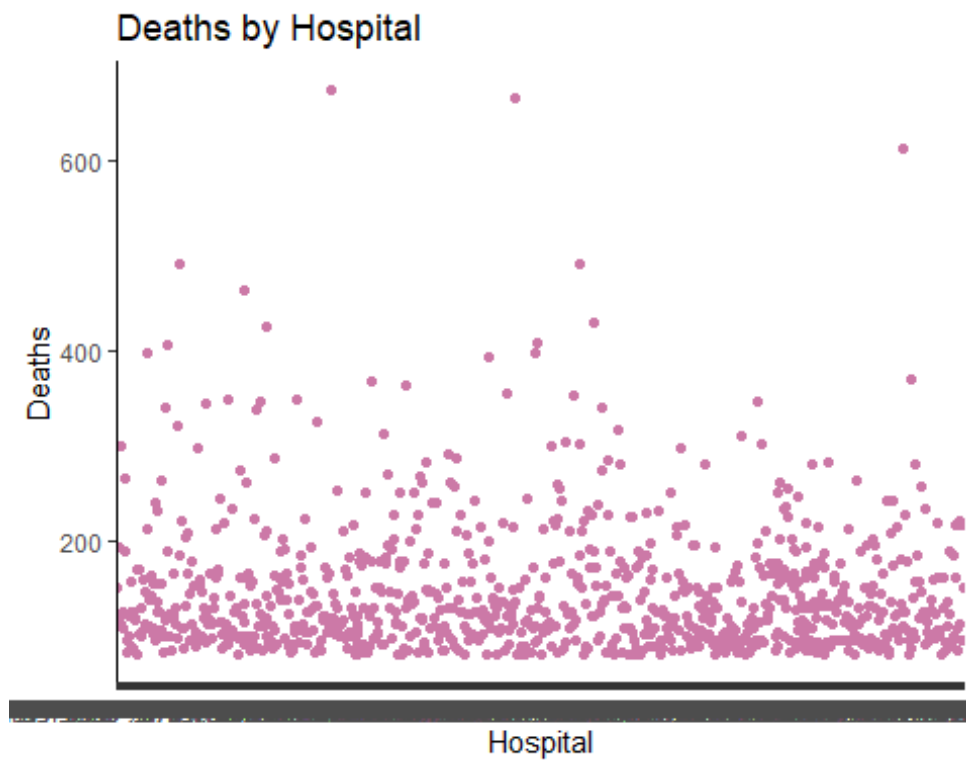


Chart 3

`bystate(col)`

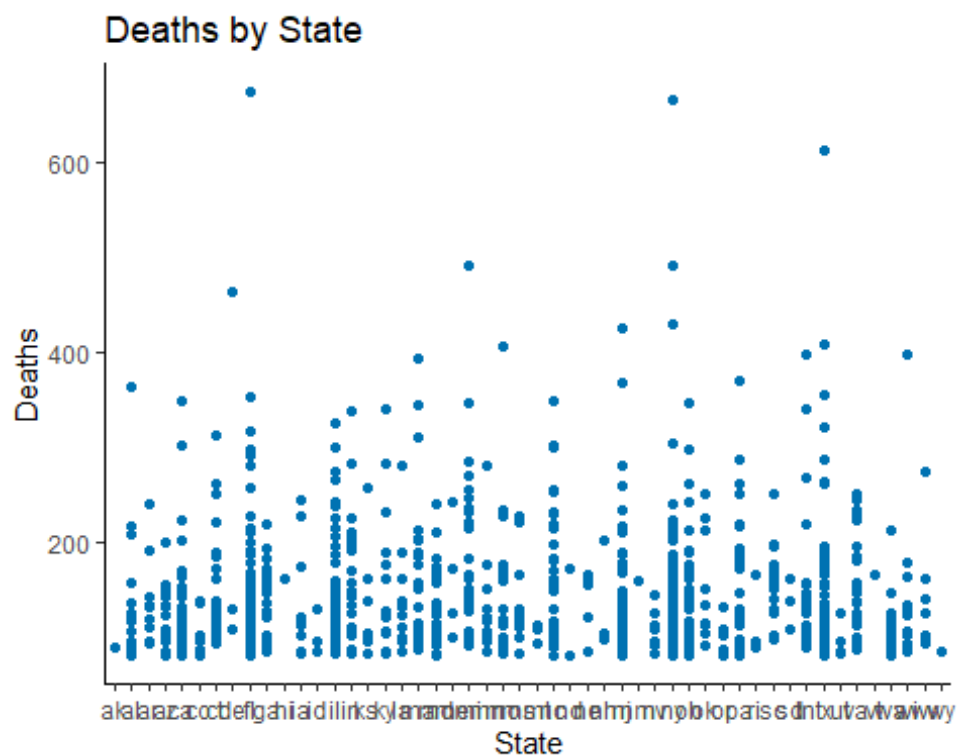
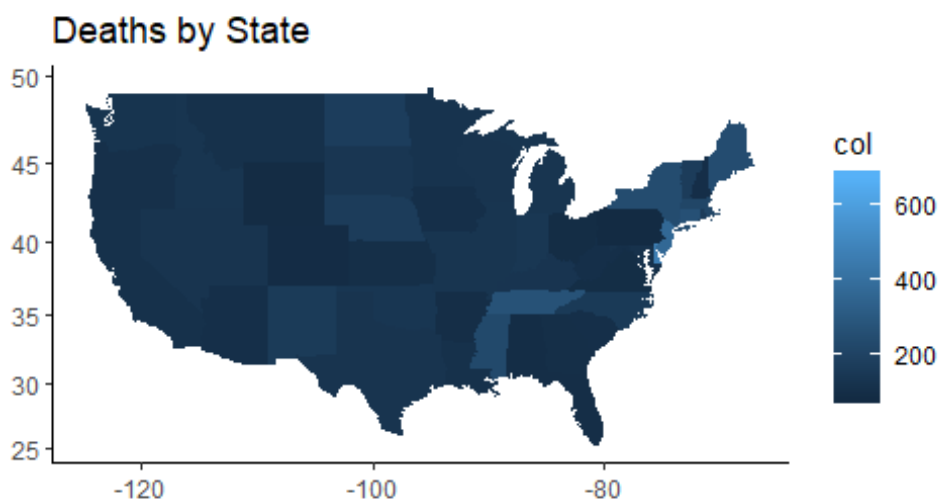


Chart 4

```
mapvariable(col)
```



% Died in Hospital

Chart 1

```
#generate descriptive statistics and histograms for each variable
col <- DAPCancerEvents$`% died in hospital`
title <- "% Died in Hospital"
```

```
title1 <- "Percent Died in Hospital by Hospital"
title2 <- "Percent Died in Hospital by State"
descriptive(DAPCancerEvents6)

## [1] "Mean: 30.4624060150376"
## [1] "Median: 29.9"
## [1] "Standard Deviation: 8.34713937287759"
## [1] "Skewness: 0.402845575498664"
## [1] "Kurtosis: 3.10253344355045"
```

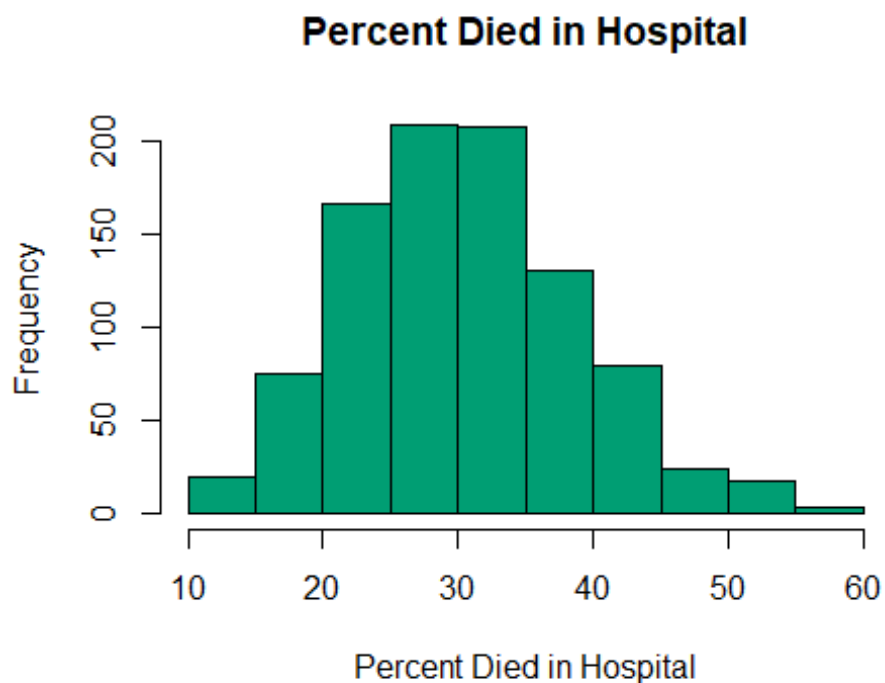


Chart 2

```
byhospital(col)
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```

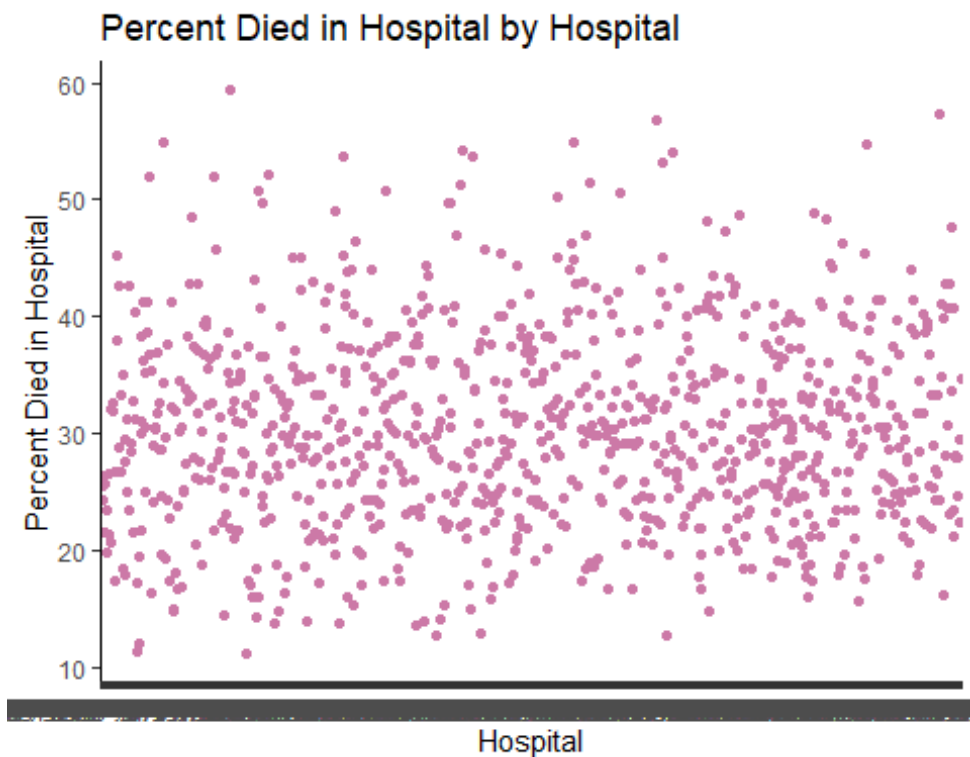


Chart 3

```
bystate(col)
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```

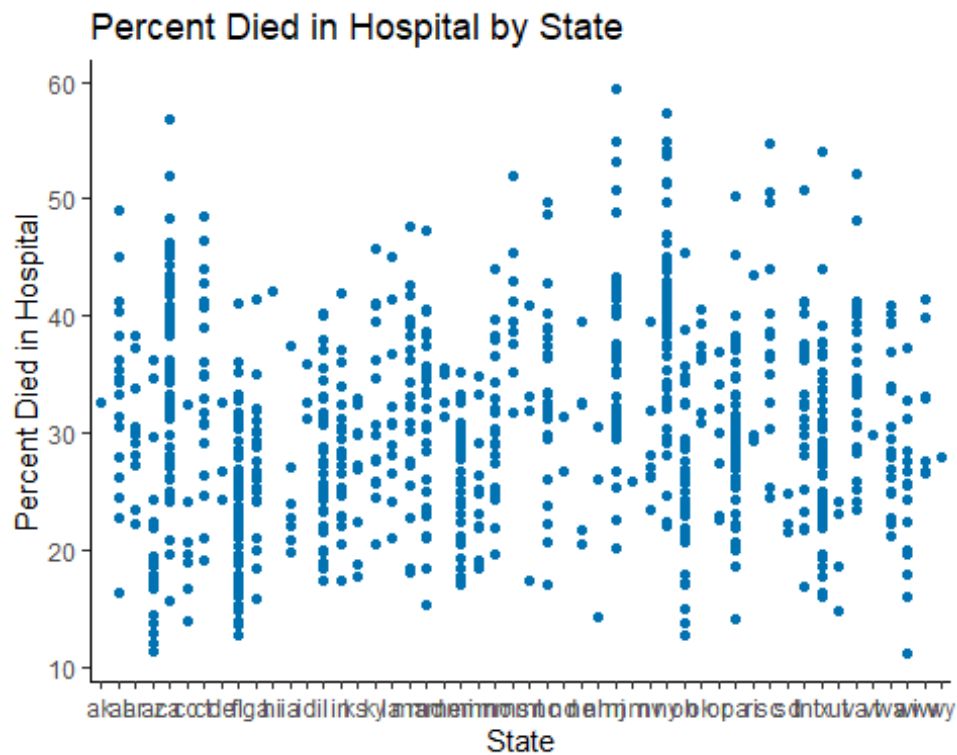
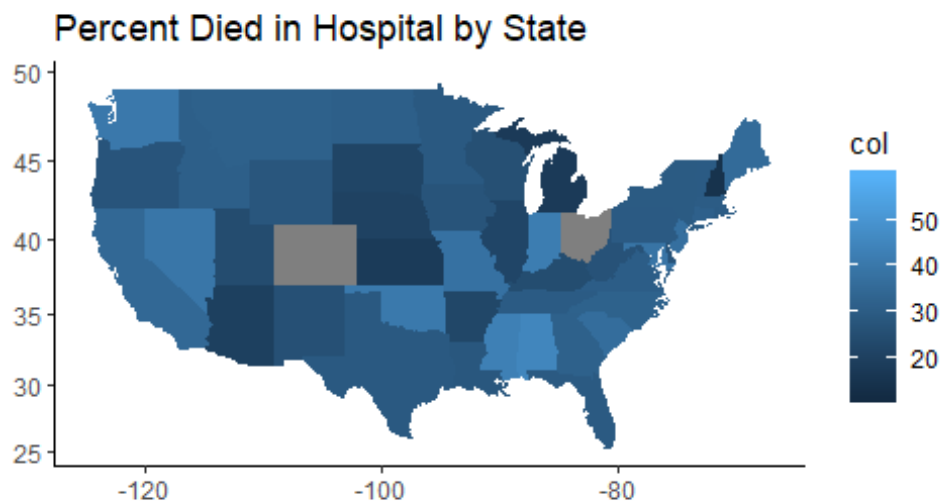


Chart 4

```
mapvariable(col)
```



% Admitted Last Month of Life

Chart 1

```
#generate descriptive statistics and histograms for each variable
col <- DAPCancerEvents$`% admitted last month`
title <- "Percent Admitted to Hospital in Last Month of Life"
title1 <- "Percent Admitted to Hospital in Last Month of Life - by Hospital"
title2 <- "Percent Admitted to Hospital in Last Month of Life - by State"
descriptive(DAPCancerEvents7)

## [1] "Mean: 64.3534115138593"
## [1] "Median: 64.2"
## [1] "Standard Deviation: 5.77795963912996"
## [1] "Skewness: 0.0179014763249724"
## [1] "Kurtosis: 3.3297094632259"
```

Percent Admitted to Hospital in Last Month of Life

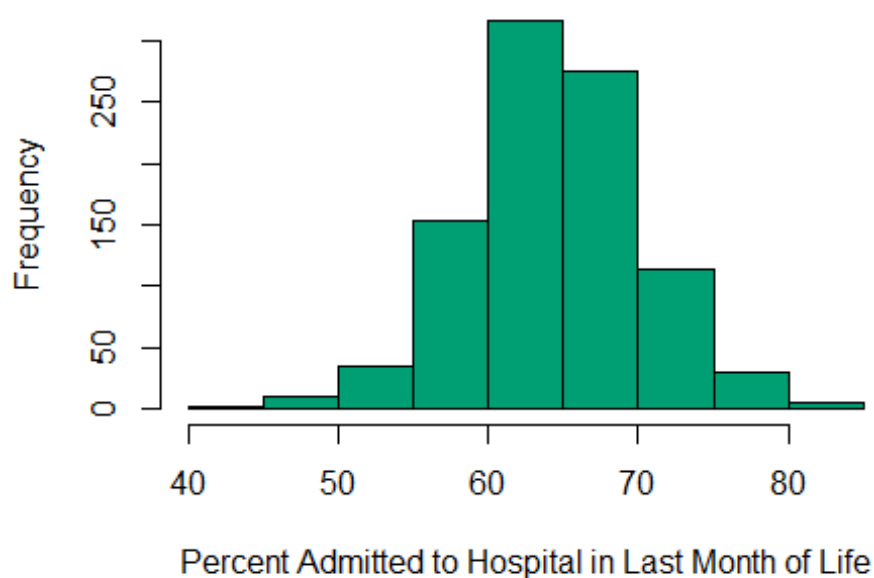


Chart 2

```
byhospital(col)
```

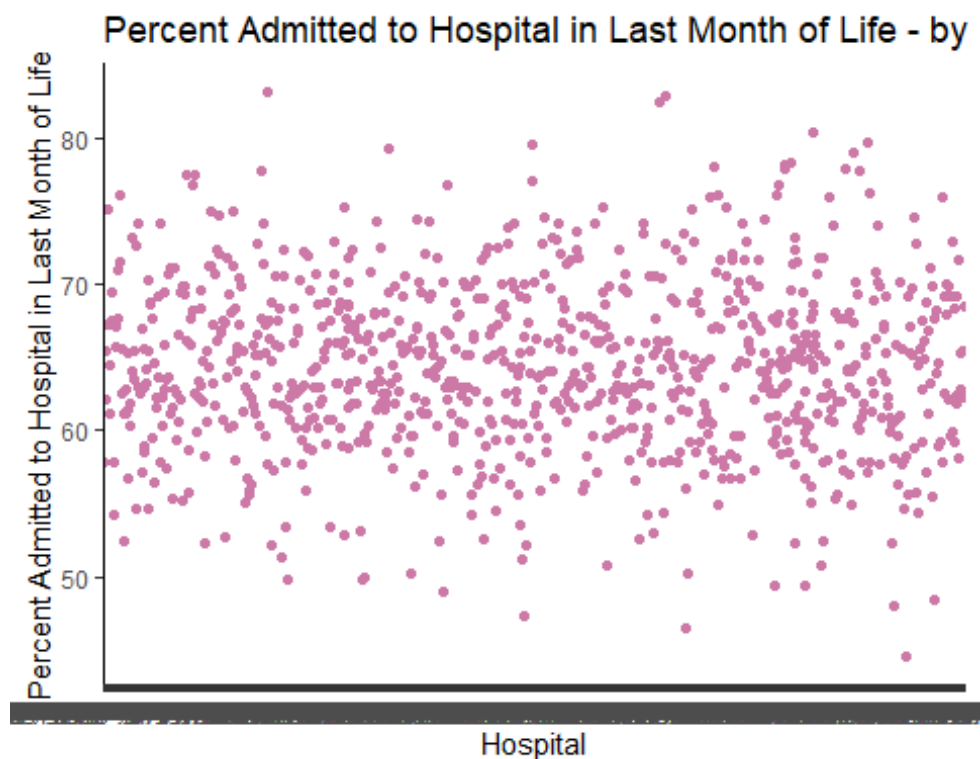


Chart 3

```
bystate(col)
```

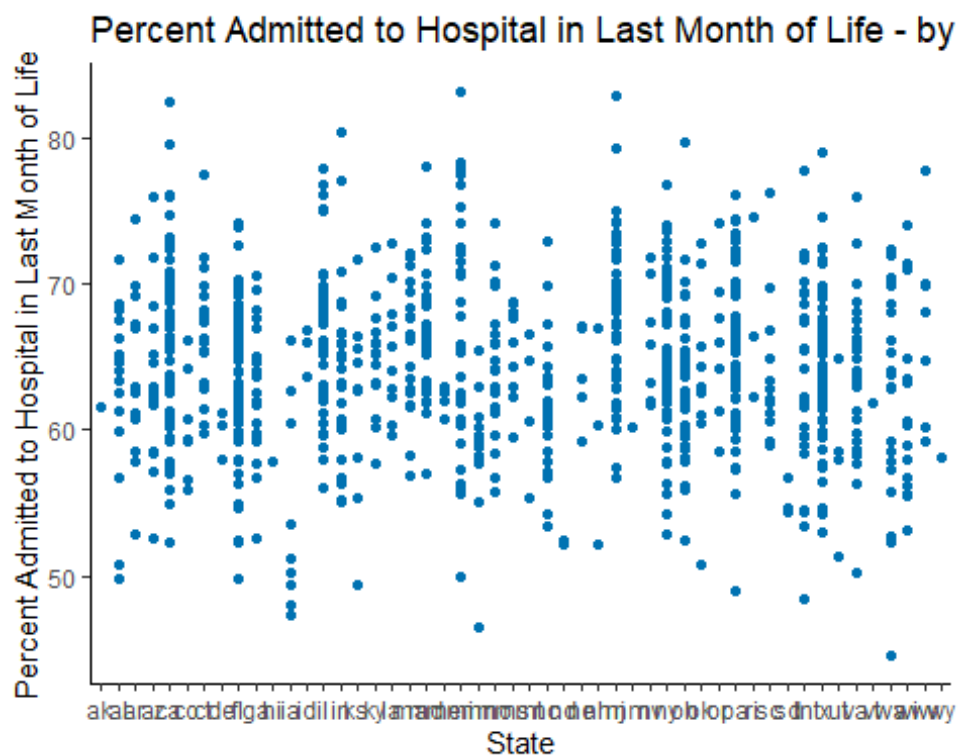
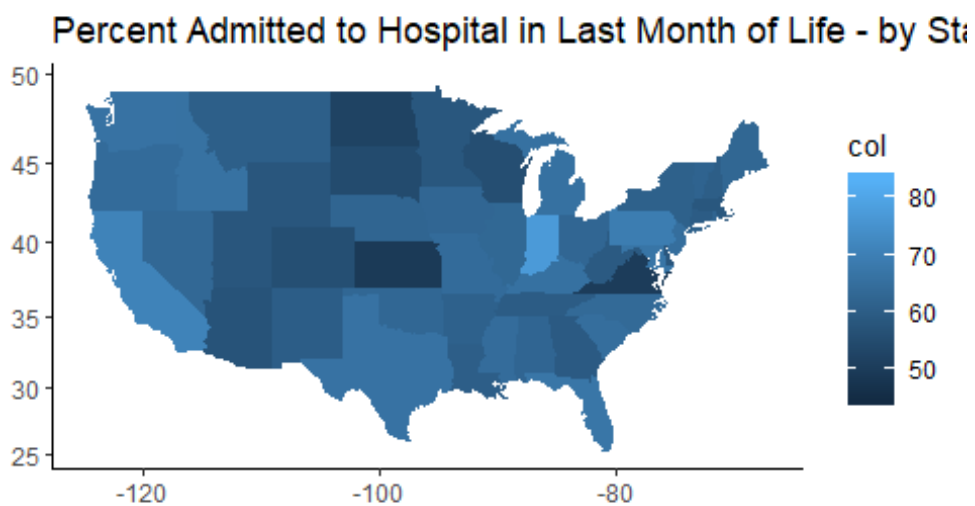


Chart 4

```
mapvariable(col)
```



Hospital Days Last Month of Life

Chart 1

```
#generate descriptive statistics and histograms for each variable
col <- DAPCancerEvents$`Hospital days last month`
title <- "Hospital Days Last Month of Life"
```

```
title1 <- "Hospital Days Last Month of Life - by Hospital"
title2 <- "Hospital Days Last Month - by State"
descriptive(DAPCancerEvents8)

## [1] "Mean: 5.74456289978678"
## [1] "Median: 5.7"
## [1] "Standard Deviation: 0.976914572494382"
## [1] "Skewness: 0.295076647718844"
## [1] "Kurtosis: 3.21604510459635"
```

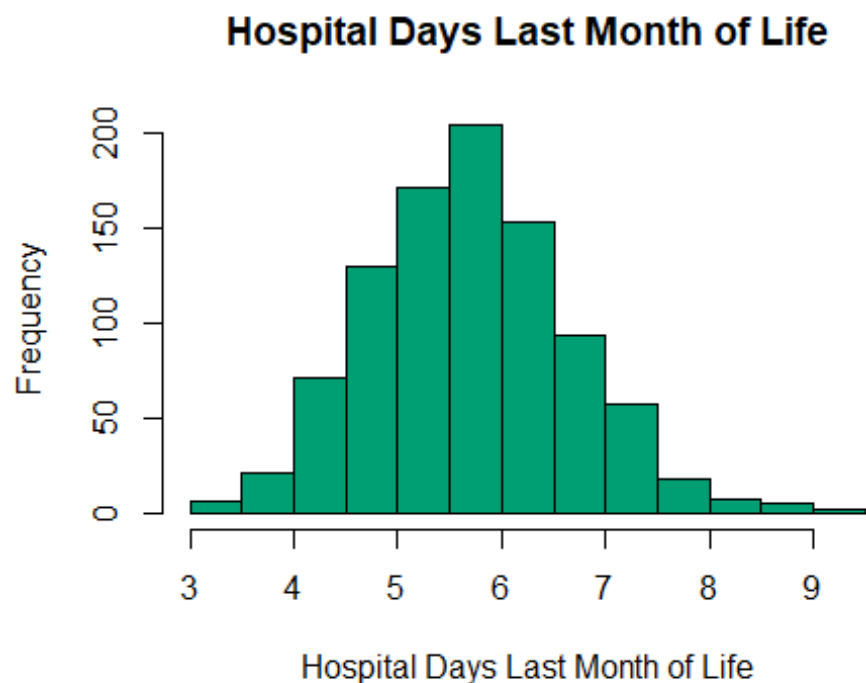


Chart 2
byhospital(col)

Hospital Days Last Month of Life - by Hospital

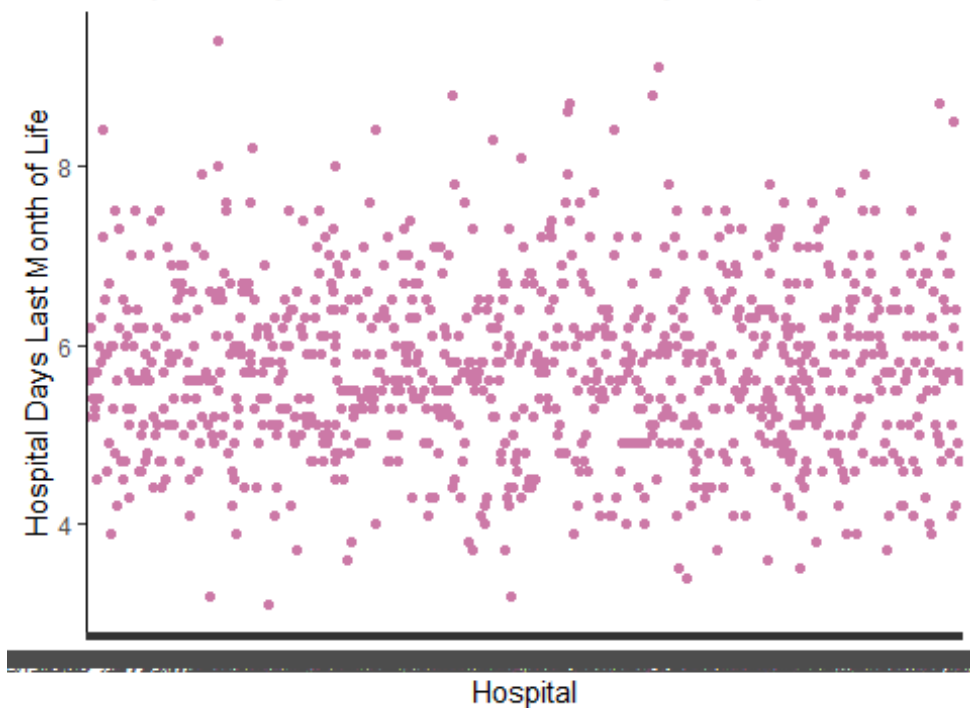


Chart 3

```
#### Chart 3
bystate(col)
```

Hospital Days Last Month - by State

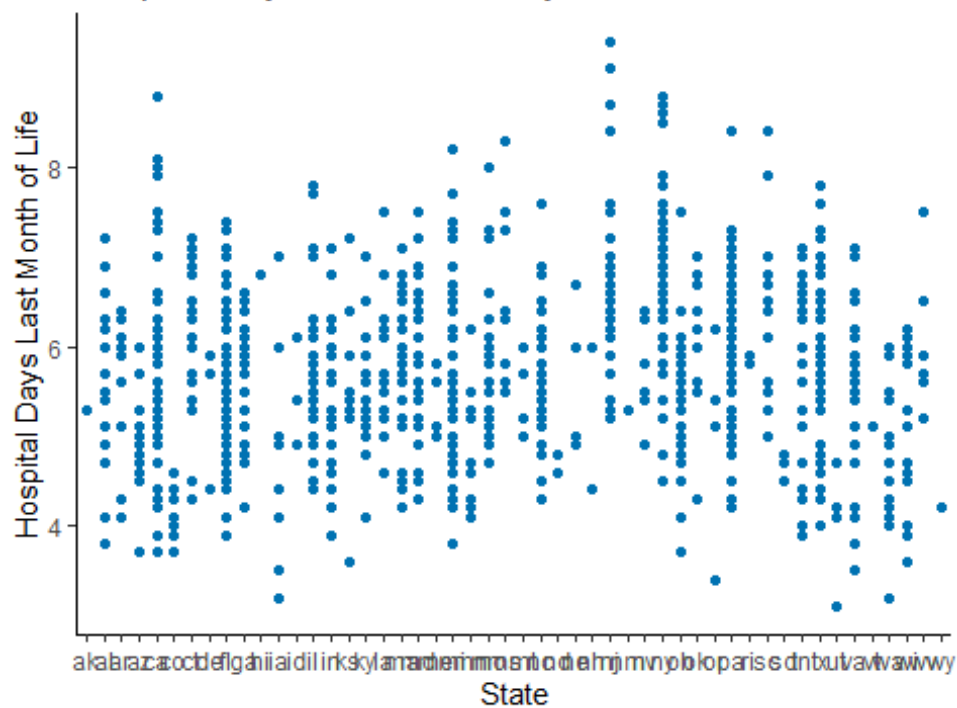
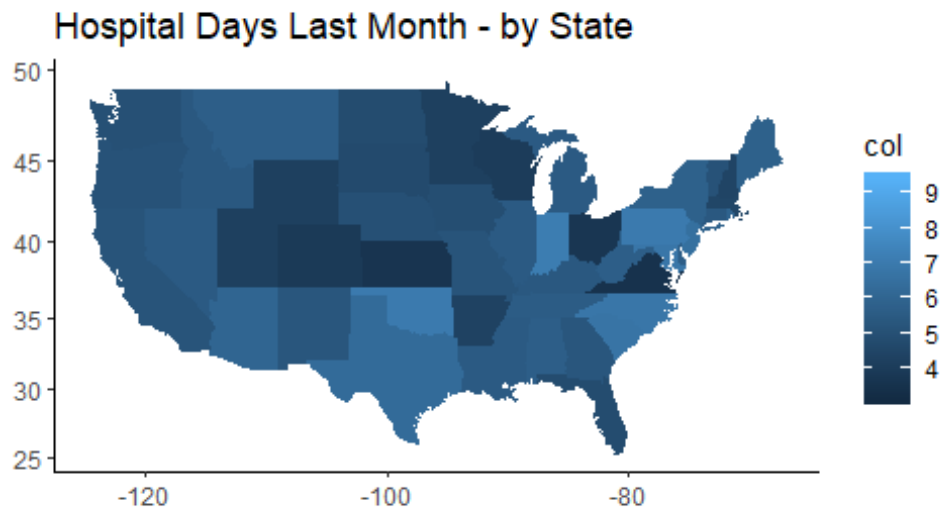


Chart 4

```
#### Chart 4
mapvariable(col)
```



% Admitted to ICU Last Month of Life

Chart 1

```
#generate descriptive statistics and histograms for each variable
col <- DAPCancerEvents$`% admitted ICU last month`
title <- "Percent Admitted to ICU Last Month of Life"
title1 <- "Percent Admitted to ICU Last Month of Life - by Hospital"
title2 <- "Percent Admitted to ICU Last Month of Life - by State"
descriptive(DAPCancerEvents9)

## [1] "Mean: 25.6288209606987"
## [1] "Median: 24.4"
## [1] "Standard Deviation: 9.36851949431984"
## [1] "Skewness: 0.688880827493528"
## [1] "Kurtosis: 3.28072116787985"
```

Percent Admitted to ICU Last Month of Life

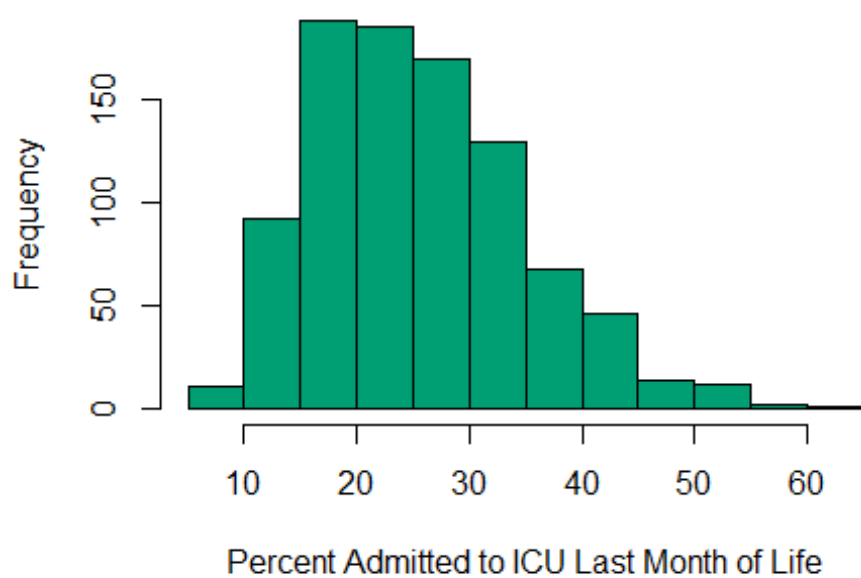


Chart 2

```
byhospital(col)
```

```
## Warning: Removed 22 rows containing missing values (geom_point).
```

Percent Admitted to ICU Last Month of Life - by Hospital

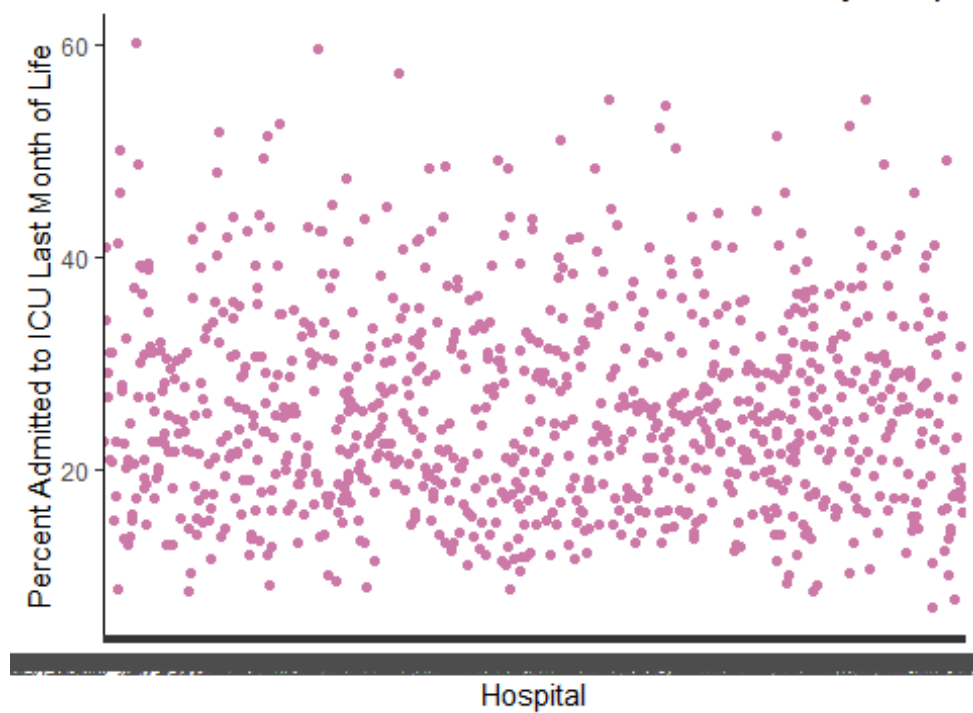


Chart 3

```
bystate(col)
```

```
## Warning: Removed 22 rows containing missing values (geom_point).
```

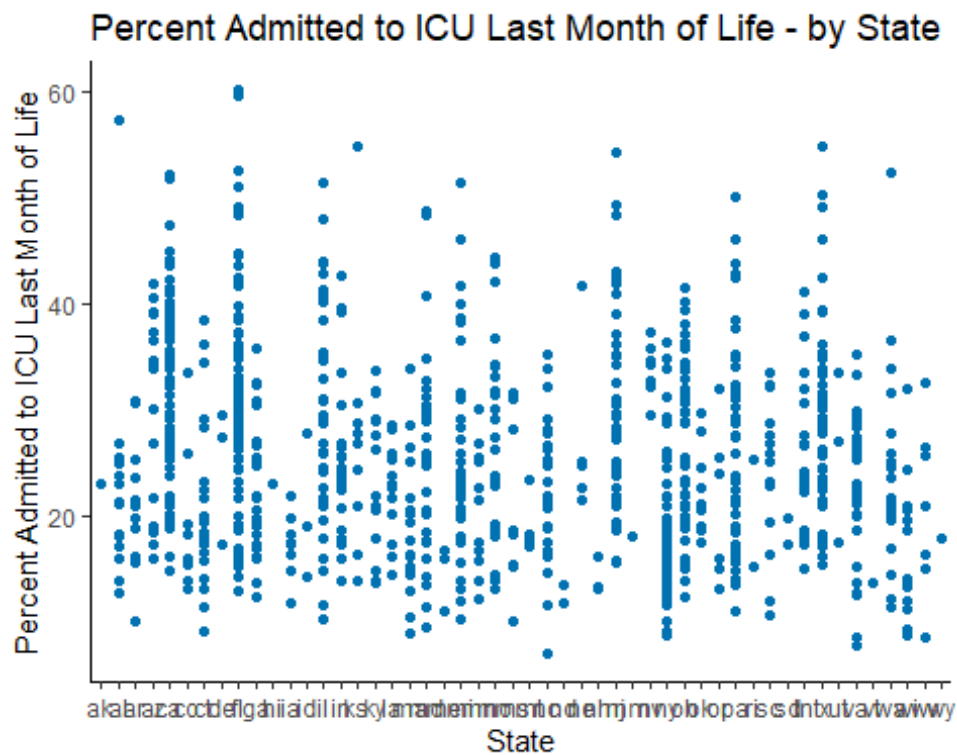
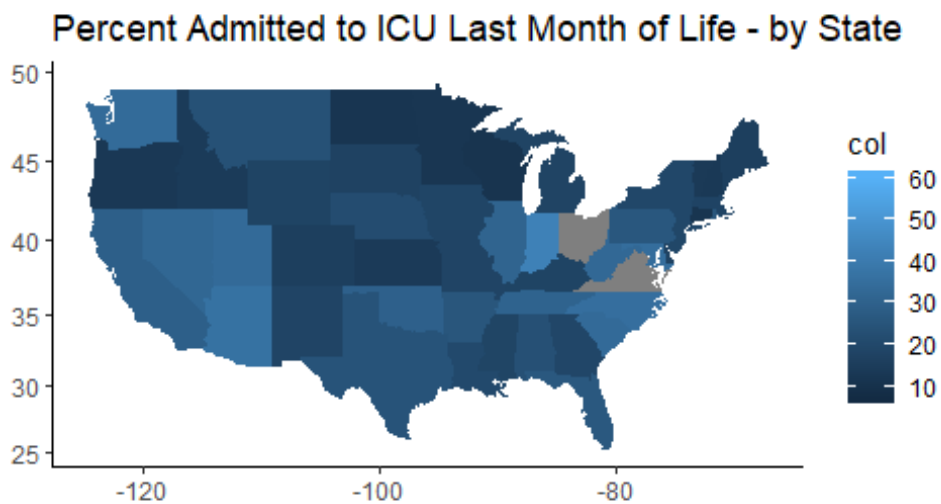


Chart 4

```
mapvariable(col)
```



ICU Days Last Month of Life

Chart 1

```
#generate descriptive statistics and histograms for each variable
col <- DAPCancerEvents$`ICU days last month`
title <- "ICU Days Last Month of Life"
```

```

title1 <- "ICU Days Last Month of Life - by Hospital"
title2 <- "ICU Days Last Month of Life - by State"
descriptive(DAPCancerEvents10)

## [1] "Mean: 1.28296943231441"
## [1] "Median: 1.2"
## [1] "Standard Deviation: 0.768060735263603"
## [1] "Skewness: 1.20461150307298"
## [1] "Kurtosis: 5.19943195537825"

```

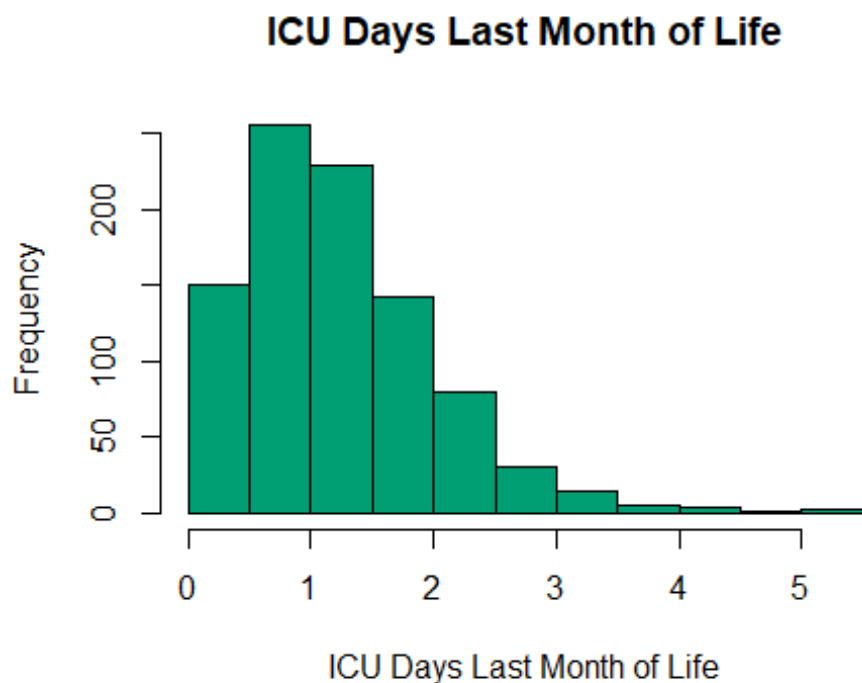


Chart 2

```
byhospital(col)
```

```
## Warning: Removed 22 rows containing missing values (geom_point).
```

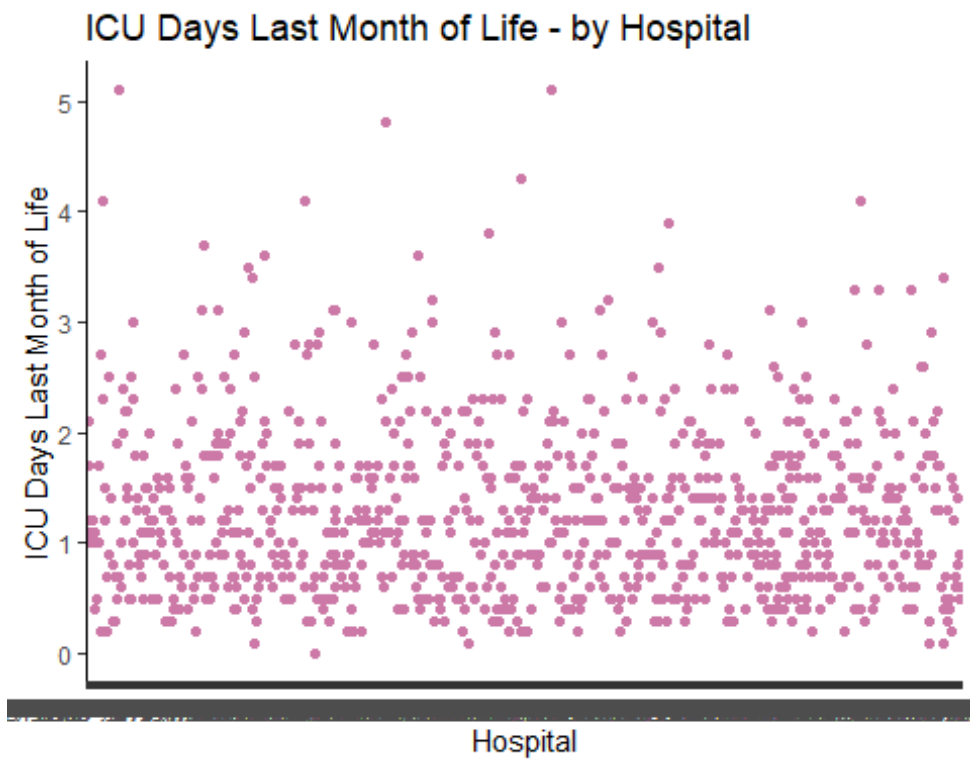


Chart 3

```
bystate(col)
```

```
## Warning: Removed 22 rows containing missing values (geom_point).
```

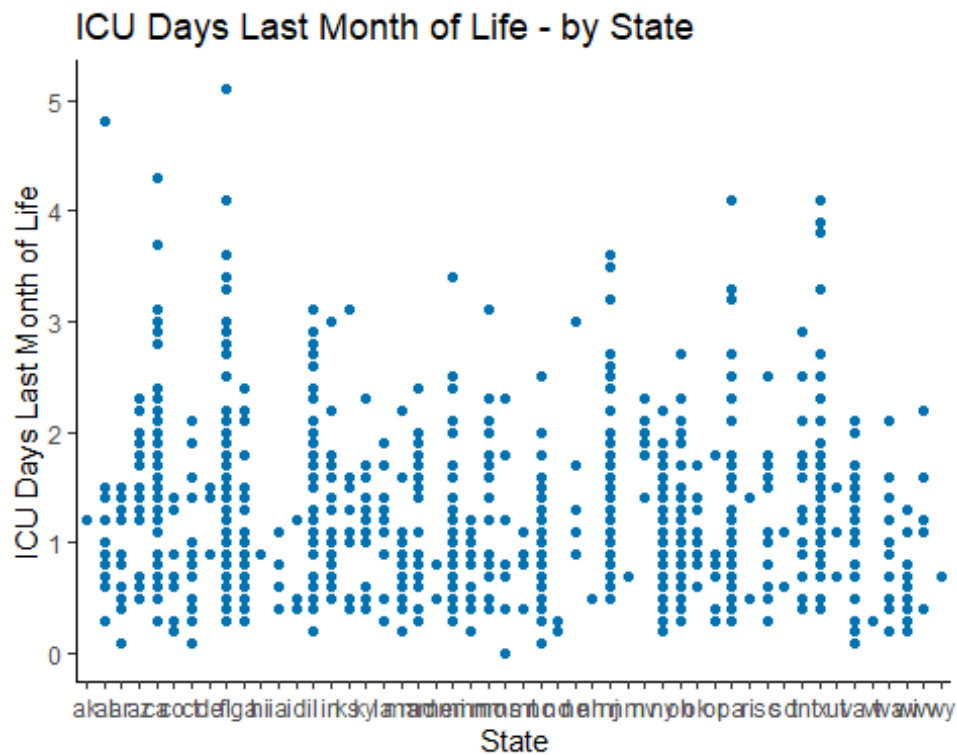
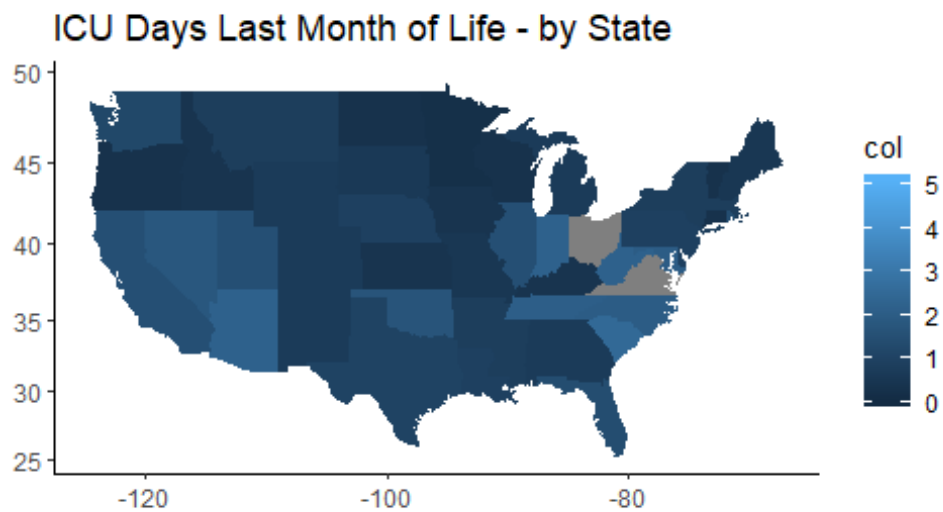


Chart 4

```
mapvariable(col)
```



% receiving life sustaining treatment last month of life

Chart 1

```
#generate descriptive statistics and histograms for each variable
col <- DAPCancerEvents$`% life sust treatment last month`
title <- "Percent Life Sustaining Treatment Last Month of Life"
title1 <- "Percent Life Sustaining Treatment Last Month of Life - by Hospital"
title2 <- "Percent Life Sustaining Treatment Last Month of Life - by State"
descriptive(DAPCancerEvents11)

## [1] "Mean: 11.7425454545455"
## [1] "Median: 11.05"
## [1] "Standard Deviation: 3.69081576530349"
## [1] "Skewness: 0.951478671995582"
## [1] "Kurtosis: 4.34101308834186"
```

Percent Life Sustaining Treatment Last Month of L

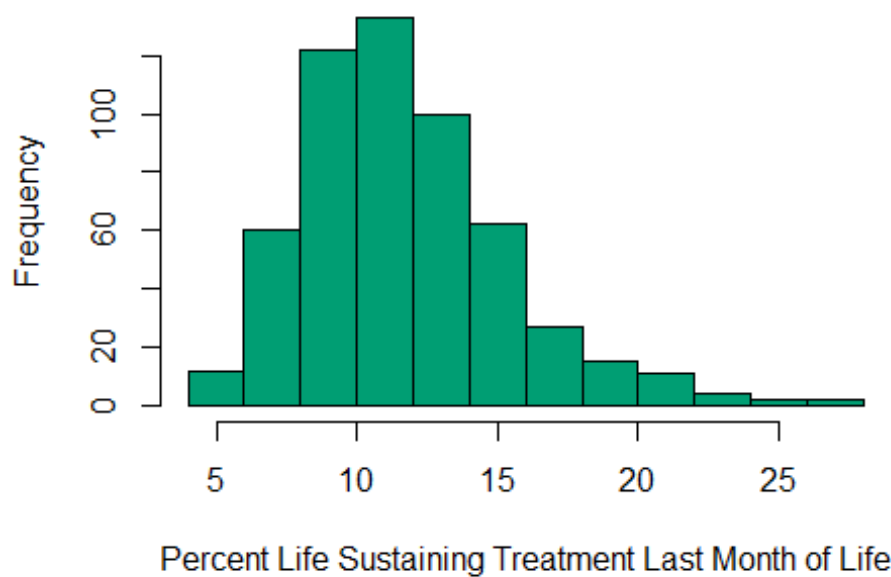


Chart 2
byhospital(col)

Warning: Removed 387 rows containing missing values (geom_point).

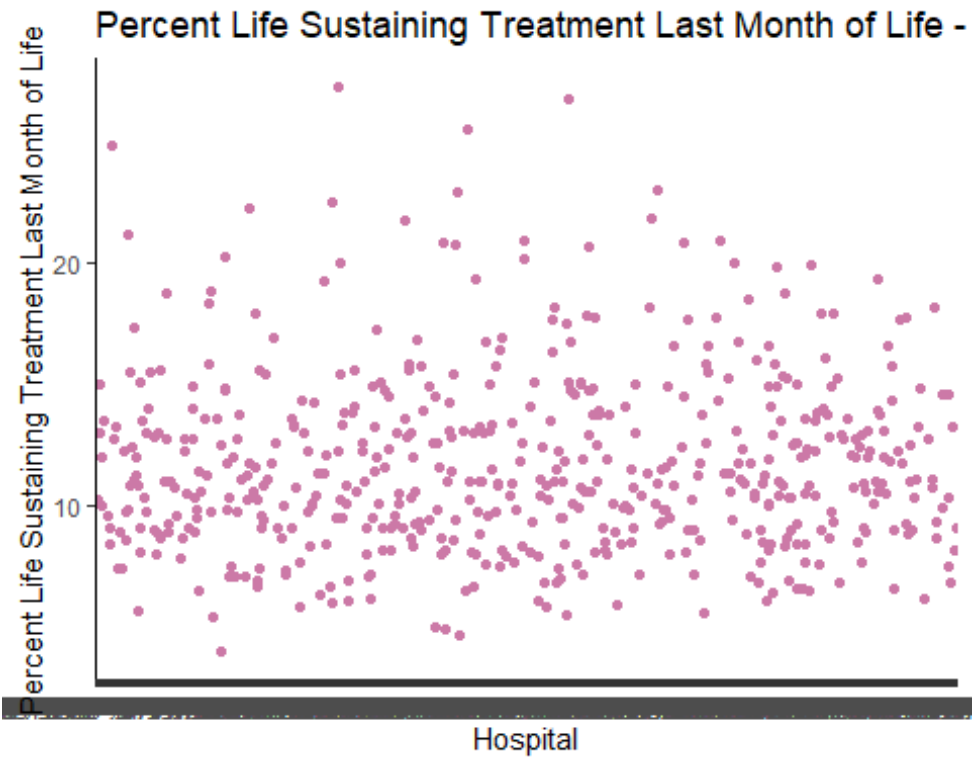


Chart 3
bystate(col)

Warning: Removed 387 rows containing missing values (geom_point).

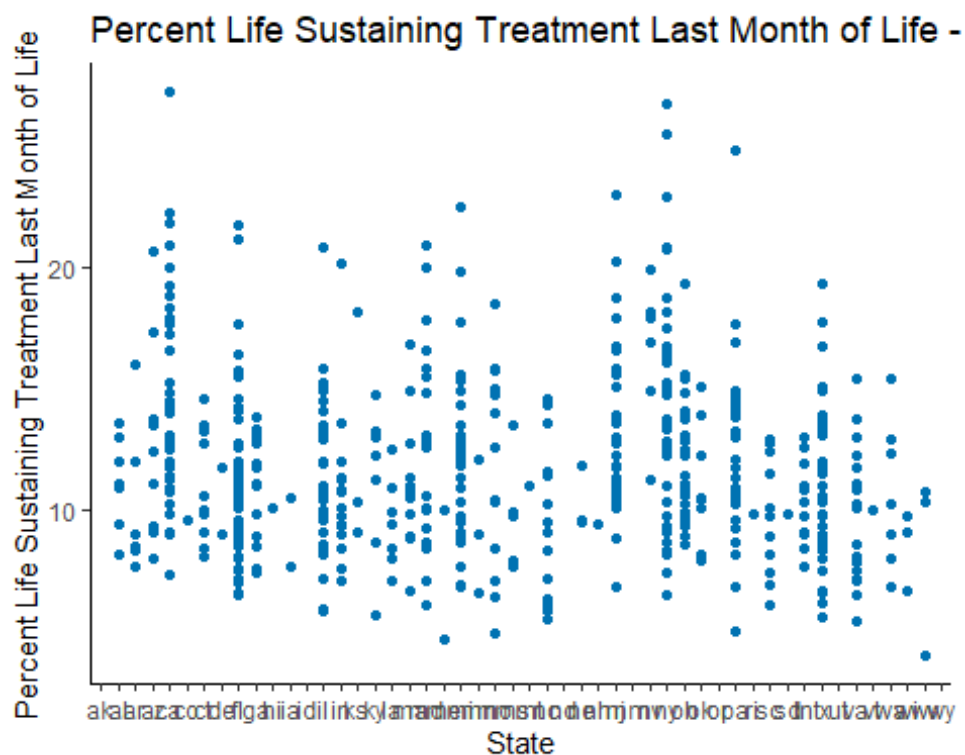
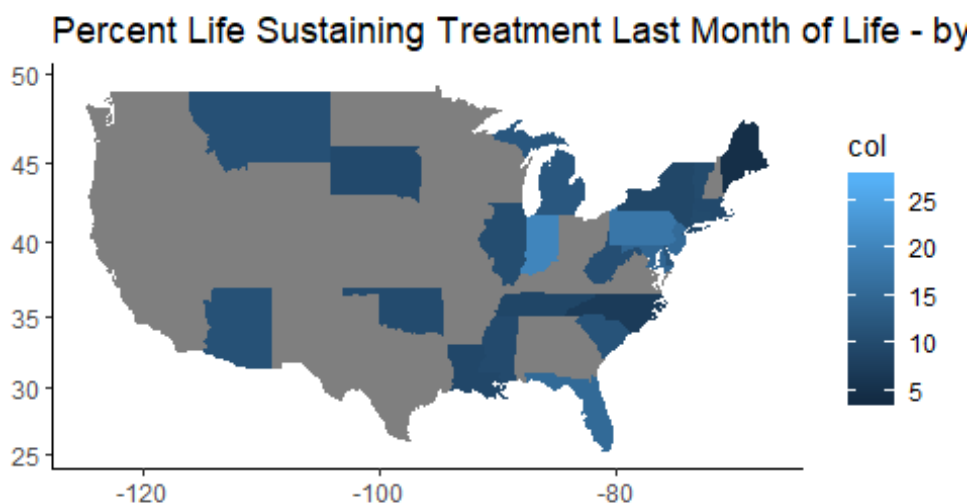


Chart 4

```
mapvariable(col)
```



% receiving chemotherapy last 2 weeks of life

Chart 1

```
#generate descriptive statistics and histograms for each variable
col <- DAPCancerEvents$`% chemo last 2 weeks`
title <- "Percent Chemo Last 2 Weeks of Life"
```

```

title1 <- "Percent Chemo Last 2 Weeks of Life - by Hospital"
title2 <- "Percent Chemo Last 2 Weeks of Life - by State"
descriptive(DAPCancerEvents12)

## [1] "Mean: 9.26740506329114"
## [1] "Median: 8.7"
## [1] "Standard Deviation: 3.42036583588304"
## [1] "Skewness: 1.24624891446332"
## [1] "Kurtosis: 5.42097748862664"

```

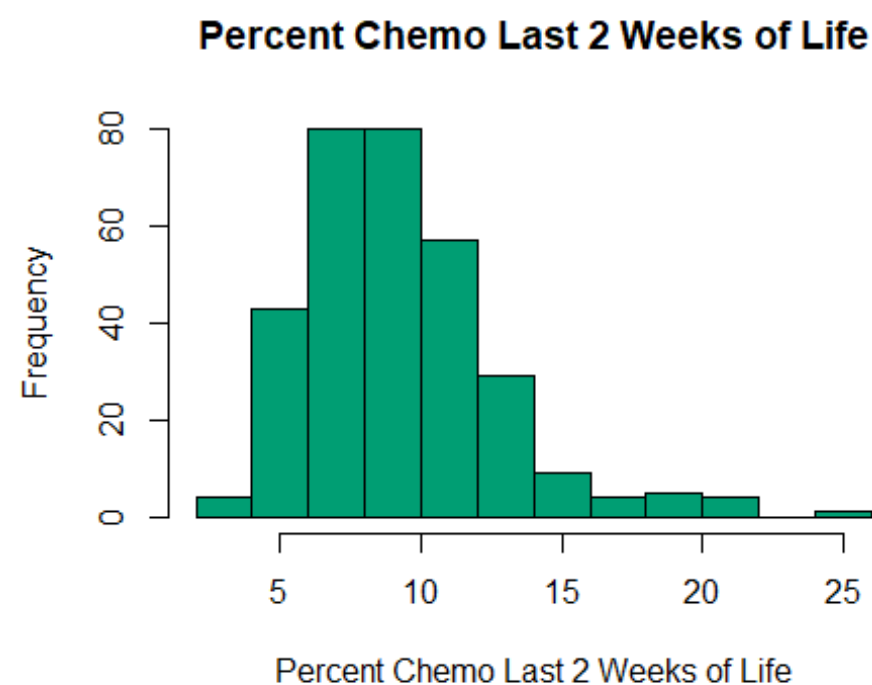


Chart 2

```

byhospital(col)

## Warning: Removed 618 rows containing missing values (geom_point).

```

Percent Chemo Last 2 Weeks of Life - by Hospital

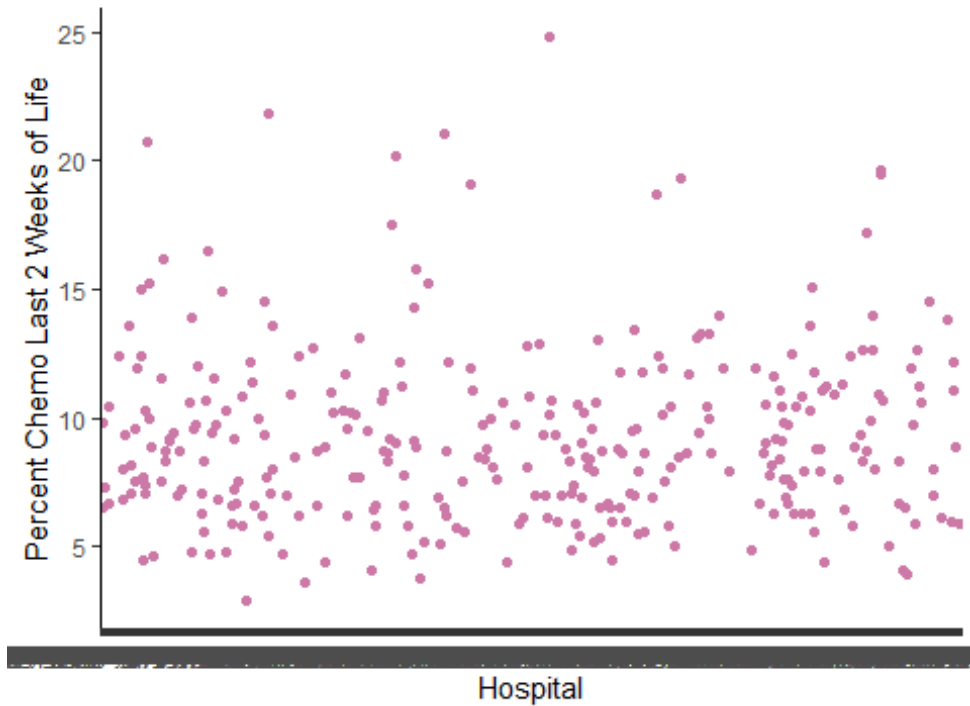


Chart 3

`bystate(col)`

Warning: Removed 618 rows containing missing values (geom_point).

Percent Chemo Last 2 Weeks of Life - by State

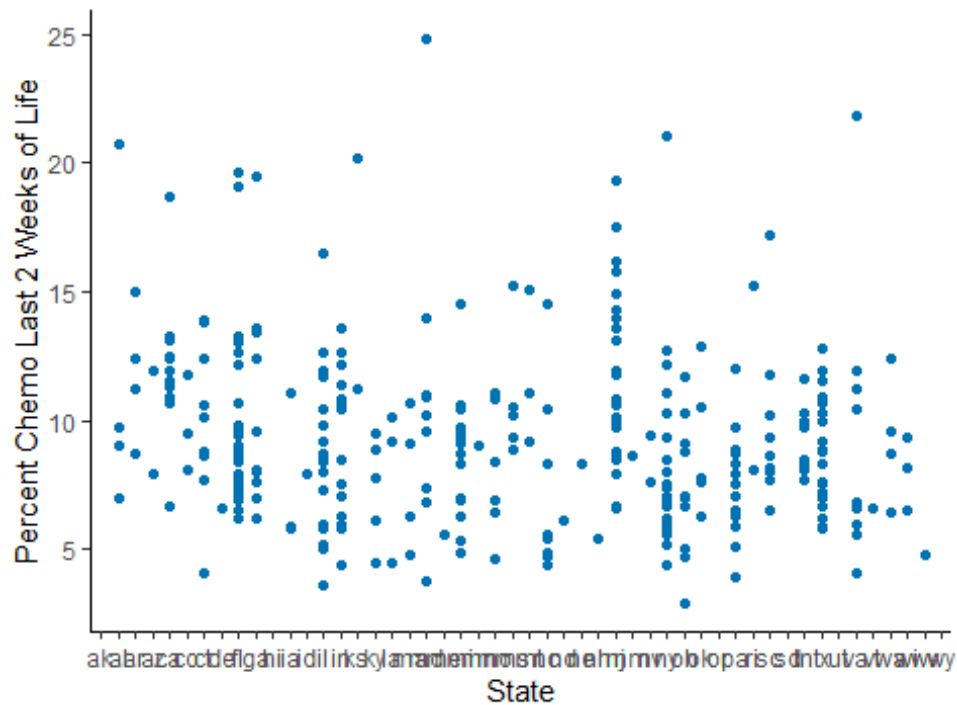
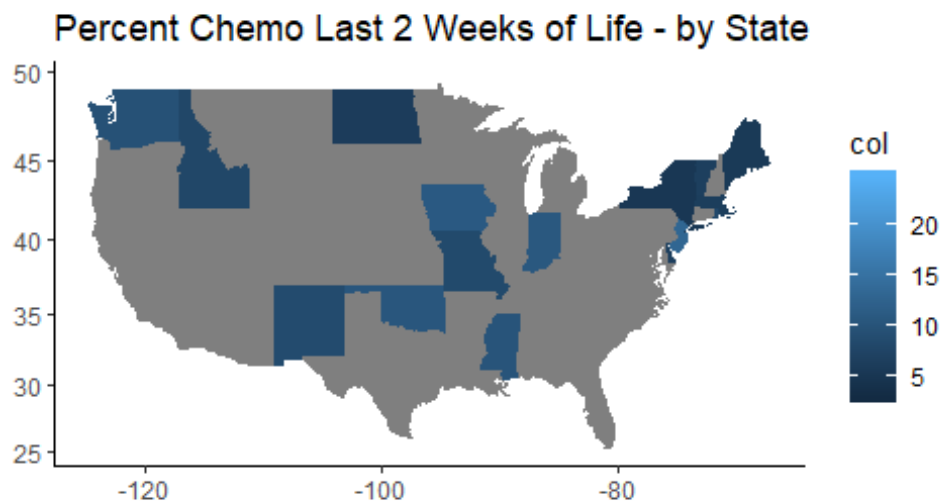


Chart 4

`mapvariable(col)`



% Hospice last month of life

Chart 1

#generate descriptive statistics and histograms for each variable

```
col <- DAPCancerEvents$`% hospice last month`
title <- "Percent Hospice Last Month of Life"
title1 <- "Percent Hospice Last Month of Life - by Hospital"
title2 <- "Percent Hospice Last Month of Life - by State"
descriptive(DAPCancerEvents13)
```

```
## [1] "Mean: 55.3044823906083"
## [1] "Median: 55.6"
## [1] "Standard Deviation: 11.8151910932693"
## [1] "Skewness: -0.256722433992258"
## [1] "Kurtosis: 3.06274209474661"
```

Percent Hospice Last Month of Life

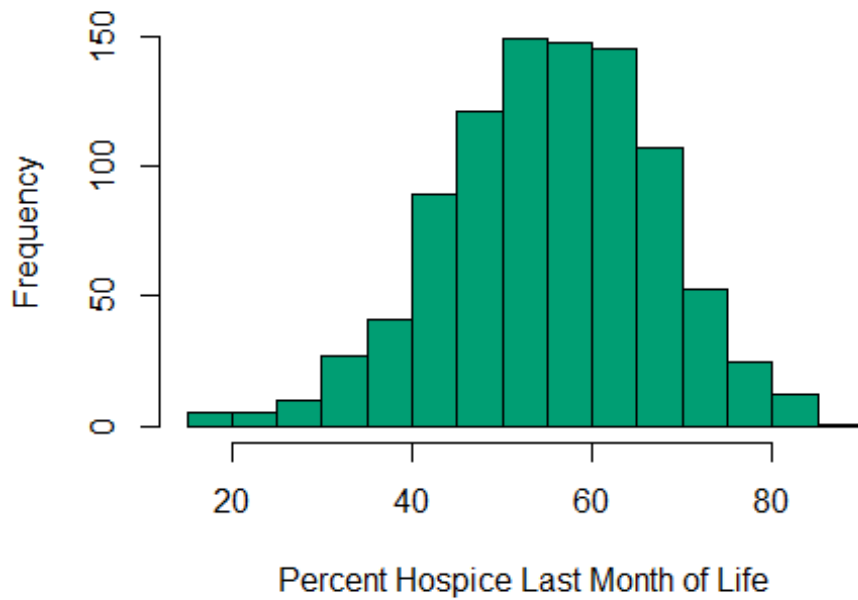


Chart 2

```
byhospital(col)
```

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

Percent Hospice Last Month of Life - by Hospital

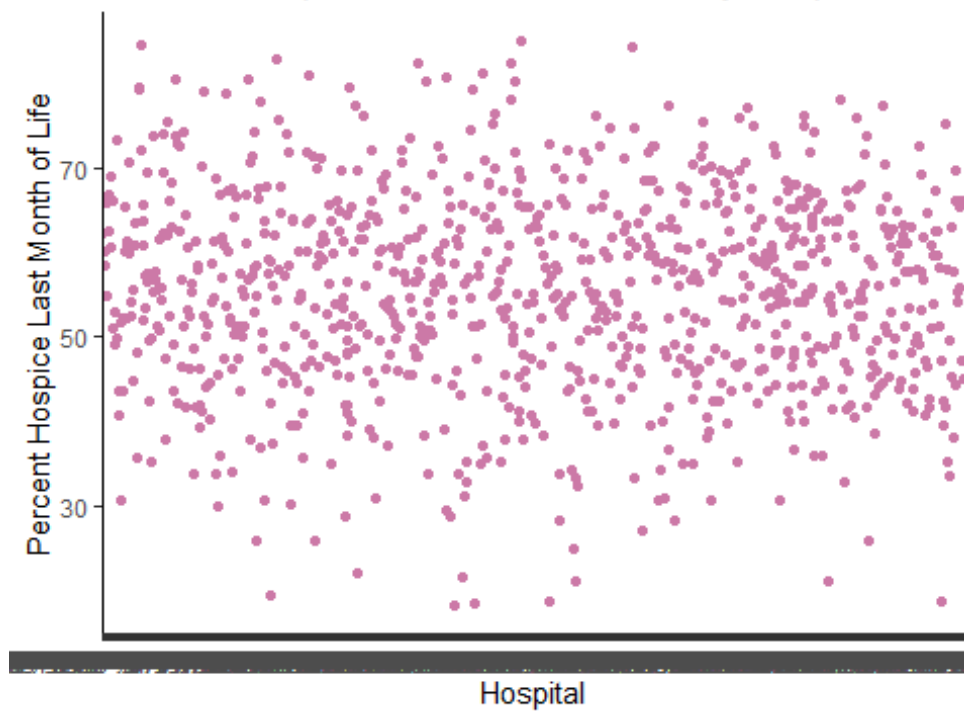


Chart 3

```
bystate(col)
```

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

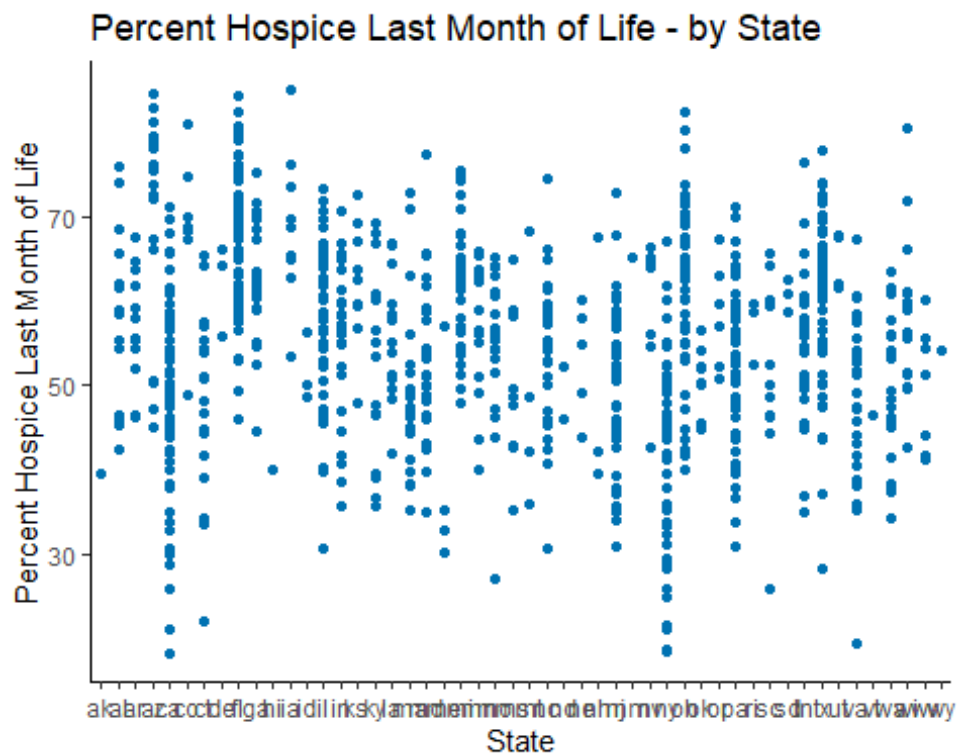
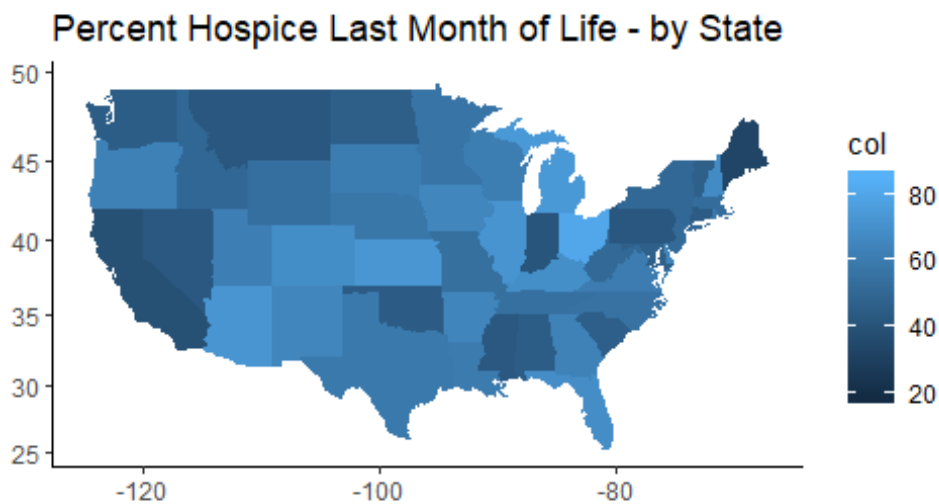


Chart 4

```
mapvariable(col)
```



Hospice Days Last Month of Life

Chart 1

```
#generate descriptive statistics and histograms for each variable
col <- DAPCancerEvents$hospice days last month`
title <- "Hospice Days Last Month of Life"
```

```

title1 <- "Hospice Days Last Month of Life - by Hospital"
title2 <- "Hospice Days Last Month of Life - by State"
descriptive(DAPCancerEvents14)

## [1] "Mean: 8.5769477054429"
## [1] "Median: 8.6"
## [1] "Standard Deviation: 2.10675604079465"
## [1] "Skewness: -0.00447642610845523"
## [1] "Kurtosis: 2.94785597006637"

```

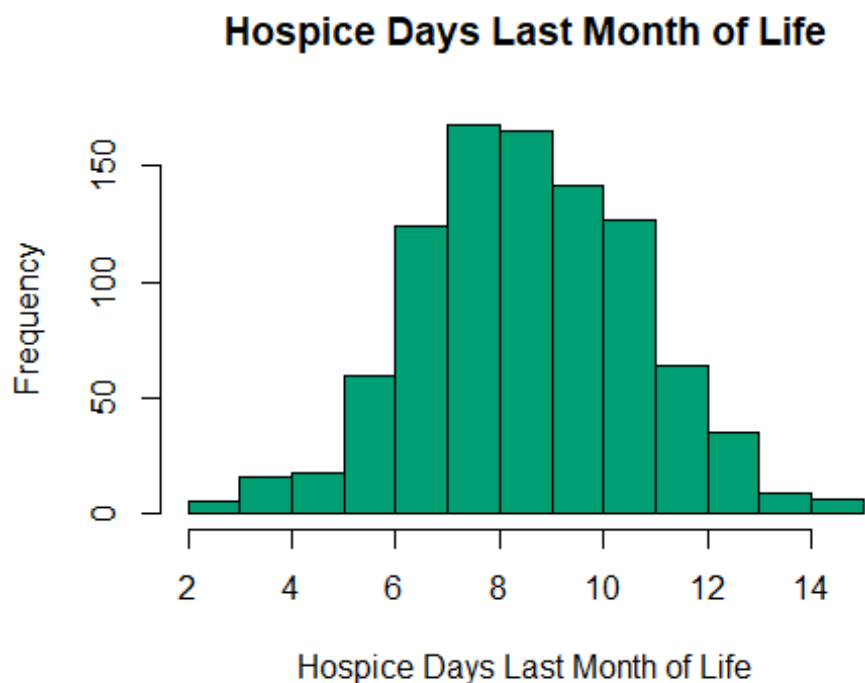


Chart 2

```
byhospital(col)
```

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

Hospice Days Last Month of Life - by Hospital

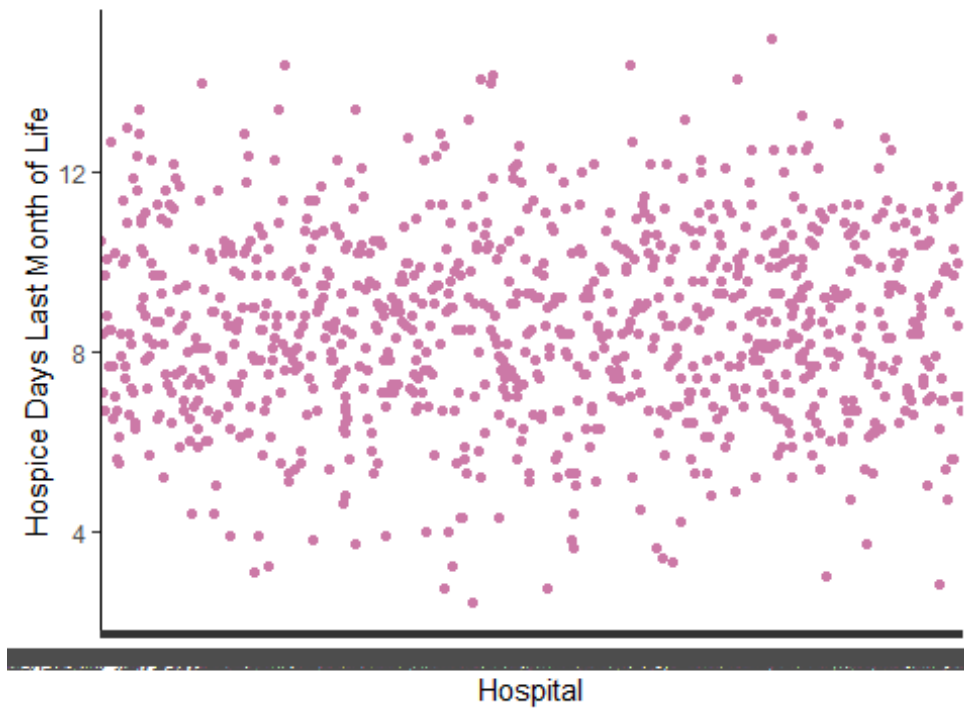


Chart 3

```
bystate(col)
```

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

Hospice Days Last Month of Life - by State

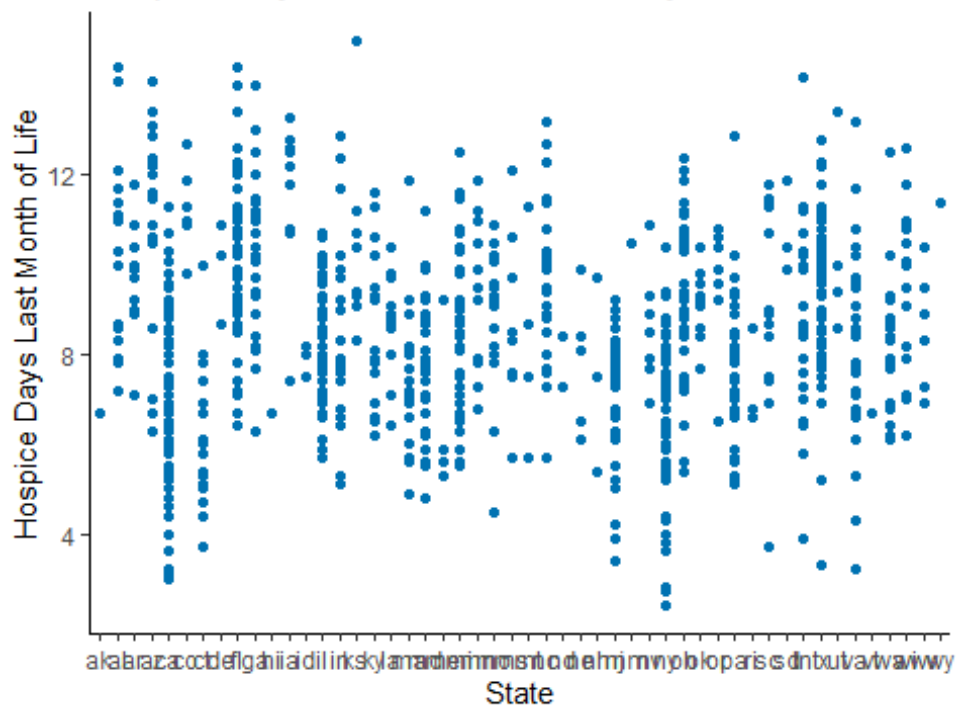
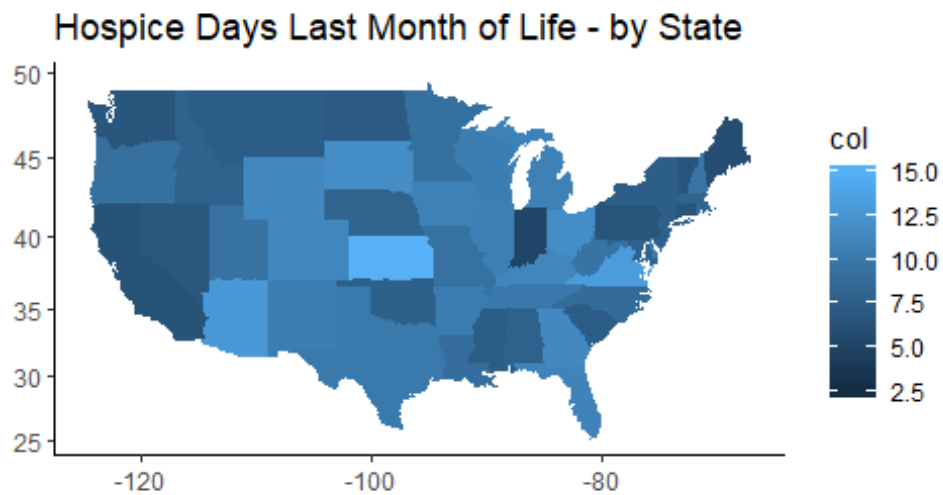


Chart 4

```
mapvariable(col)
```



% Hospice Last 3 Days of Life

Chart 1

#generate descriptive statistics and histograms for each variable

```
col <- DAPCancerEvents$`% hospice last 3 days`
title <- "% Hospice Last 3 Days of Life"
title1 <- "% Hospice Last 3 Days of Life - by Hospital"
title2 <- "% Hospice Last 3 Days of Life - by State"
descriptive(DAPCancerEvents15)
```

```
## [1] "Mean: 11.4885480572597"
## [1] "Median: 11.2"
## [1] "Standard Deviation: 3.79923845613285"
## [1] "Skewness: 0.718609957588257"
## [1] "Kurtosis: 4.01274240765306"
```

% Hospice Last 3 Days of Life



Chart 2

```
byhospital(col)
```

```
## Warning: Removed 446 rows containing missing values (geom_point).
```

% Hospice Last 3 Days of Life - by Hospital

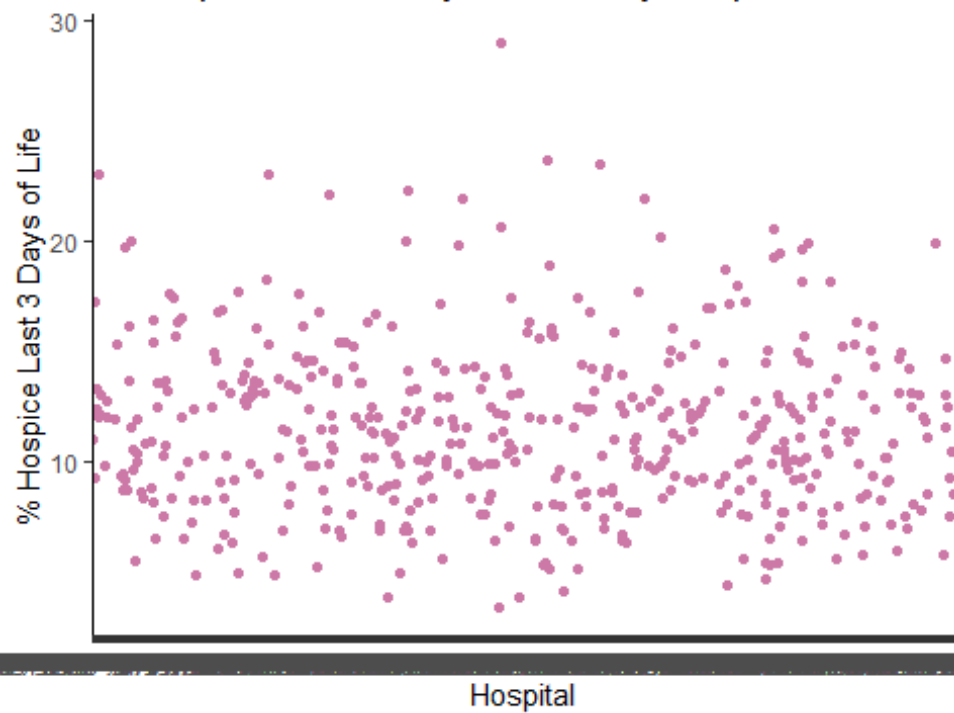


Chart 3

```
bystate(col)
```

```
## Warning: Removed 446 rows containing missing values (geom_point).
```

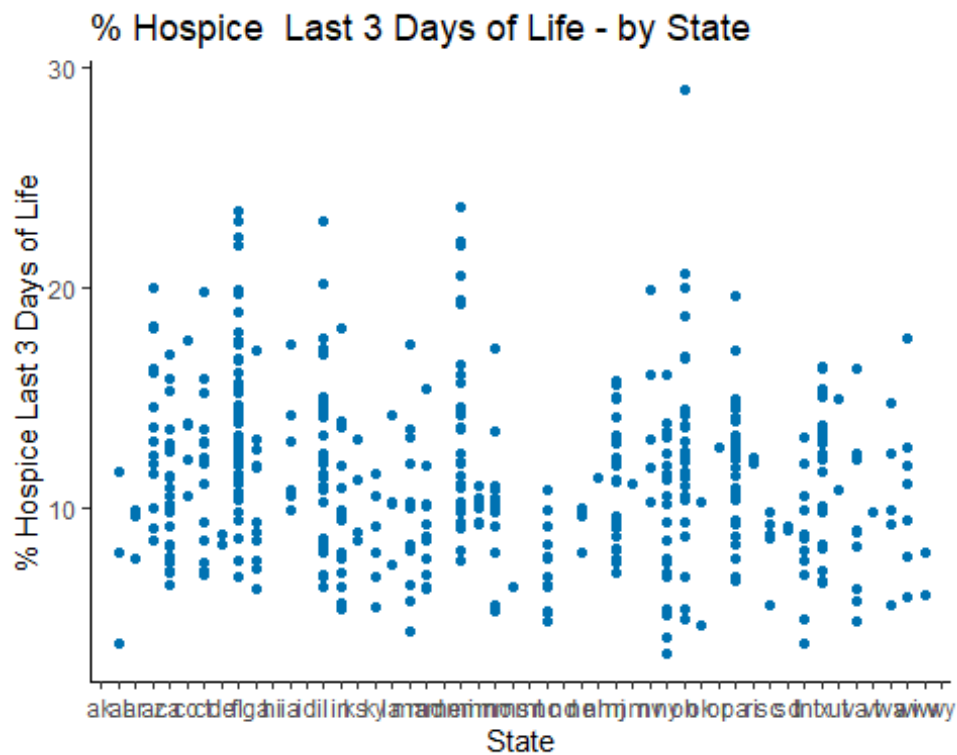
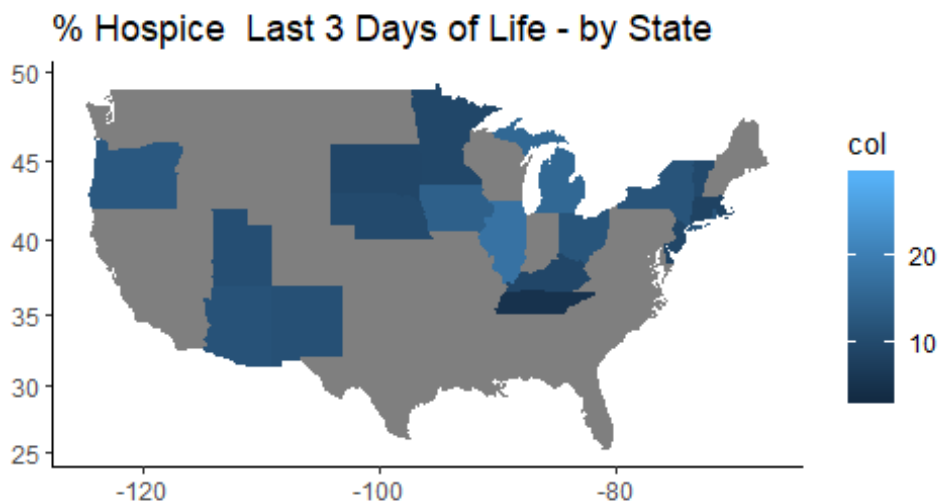


Chart 4

```
mapvariable(col)
```



% saw more than 10 physicians last 6 months of life

Chart 1

```
#generate descriptive statistics and histograms for each variable
col <- DAPCancerEvents$`% > 10 physicians last 6 months`
title <- "> 10 Physicians Seen in Last 6 Months of Life"
```

```
title1 <- "> 10 Physicians Seen in Last 6 Months of Life - by Hospital"
title2 <- "> 10 Physicians Seen in Last 6 Months of Life - by State"
descriptive(DAPCancerEvents16)

## [1] "Mean: 52.5716417910448"
## [1] "Median: 52.9"
## [1] "Standard Deviation: 12.6358211514082"
## [1] "Skewness: -0.171086564547256"
## [1] "Kurtosis: 2.62866589106455"
```

> 10 Physicians Seen in Last 6 Months of Life

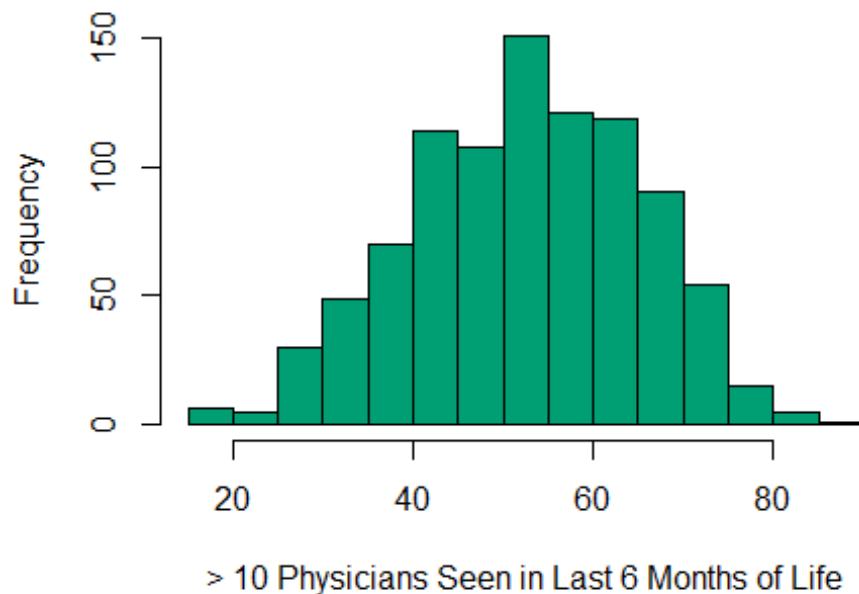


Chart 2

```
byhospital(col)
```

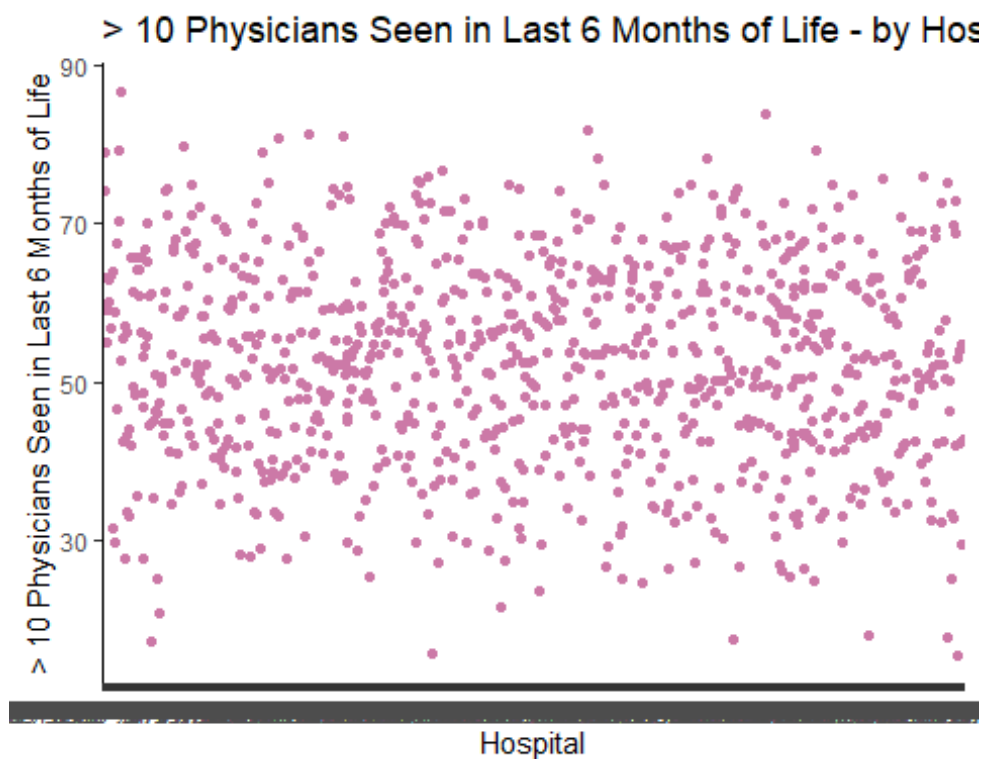


Chart 3

`bystate(col)`

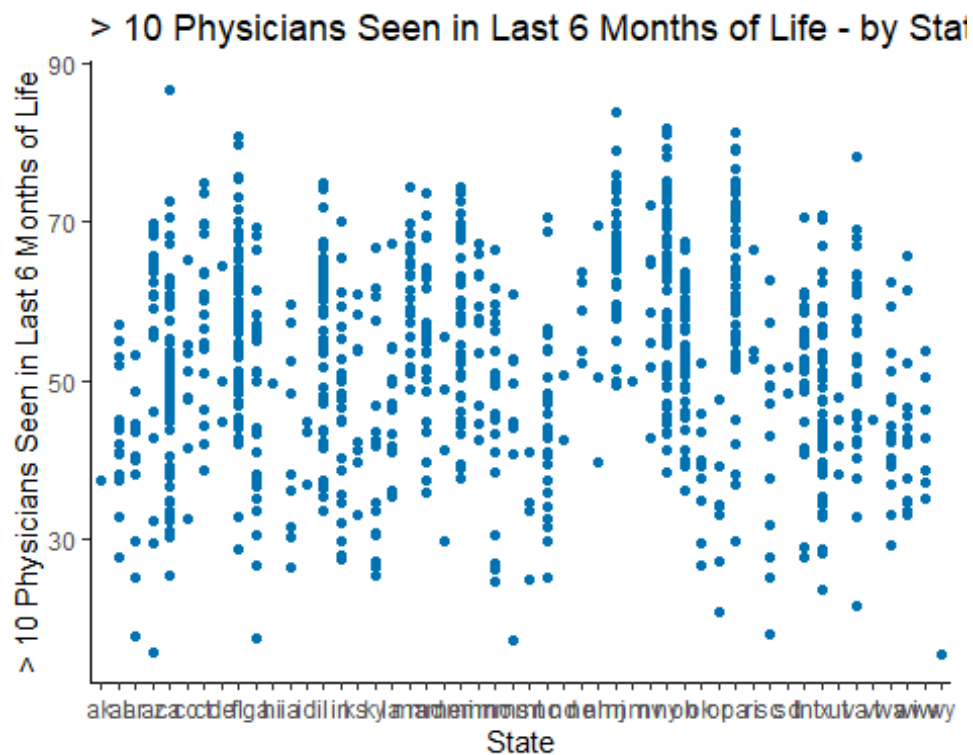
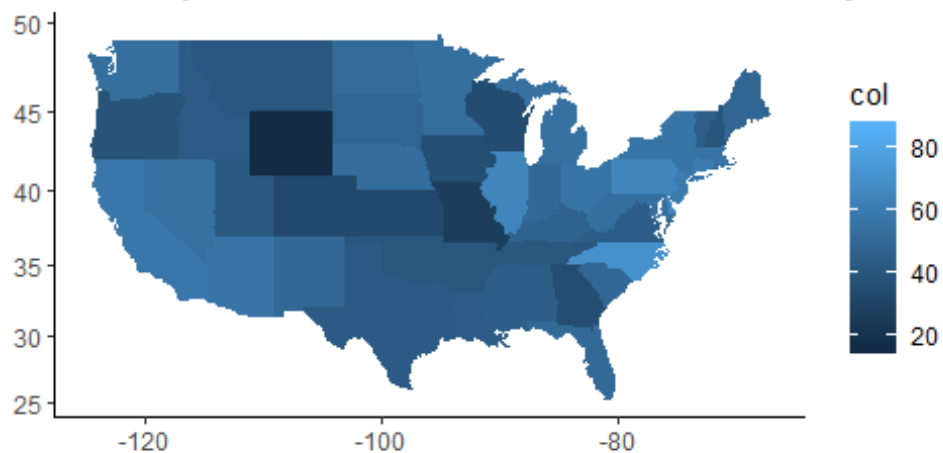


Chart 4

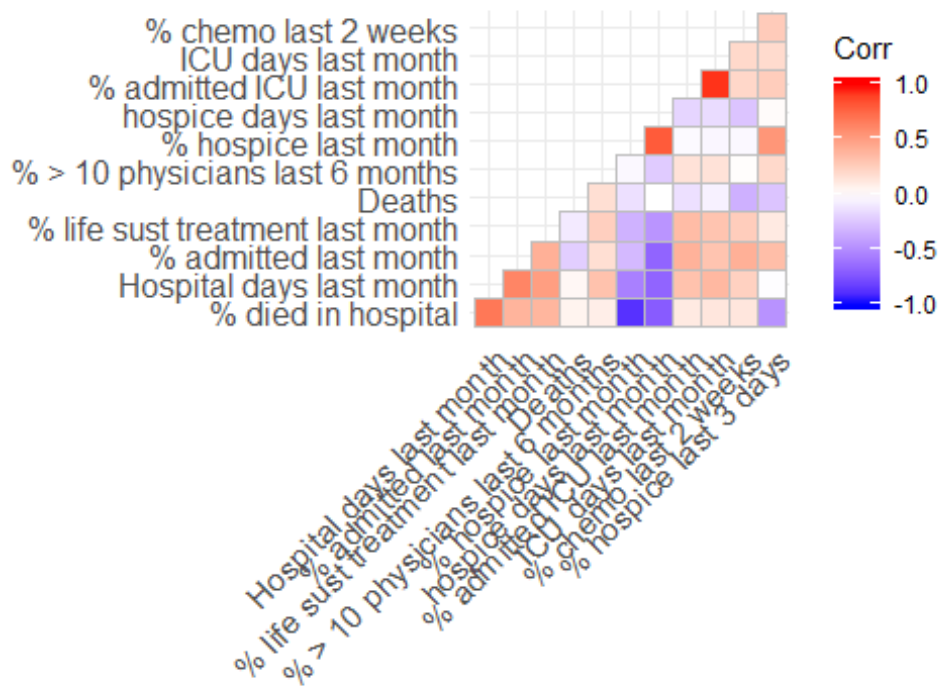
`mapvariable(col)`

> 10 Physicians Seen in Last 6 Months of Life - by State



Correlation

```
#install.packages("ggcorrplot")
library("ggcorrplot")
correlationdata <- DAPCancerEvents[,5:16]
DAP.cor<- round(cor(correlationdata,use="complete.obs"),4)
#Note used "complete.obs" rather than "everything"
#head(DAP.cor[, 1:6])
ggcorrplot(DAP.cor, hc.order=TRUE, type="lower")
```



Linear Regression

Row

predicting % hospice last month given % died in hospital

#Linear Regression

```
#regression model for predicting % hospice last month considering % died in hospital
#as the variable
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`% hospice last month`~DAPCancerEvents$`%
died in hospital`,
                        data=DAPCancerEvents)
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)
#summary(DAPCancerEvents.Lm)

paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)

## [1] "Adjusted r squared= 0.671665988785723"
```

predicting % died in hospital given number of hospice days last month

#Linear Regression

```
#regression model for predicting % died in hospital considering hospice days last month.
#as the variable
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`% died in
hospital`~DAPCancerEvents$`hospice days last month`,
                        data=DAPCancerEvents)
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)
#summary(DAPCancerEvents.Lm)
```

```
paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)
```

```
## [1] "Adjusted r squared= 0.46276897726445"
```

predicting hospital days last month given % admitted last month

```
#Linear Regression
```

```
#regression model for predicting hospital days last month considering % admitted last month.
```

```
#as the variable
```

```
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`Hospital days last month`~DAPCancerEvents$`% admitted last month`,  
                        data=DAPCancerEvents)
```

```
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)
```

```
#summary(DAPCancerEvents.lm)
```

```
paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)
```

```
## [1] "Adjusted r squared= 0.372540202579385"
```

predicting hospice days last month given % admitted last month

```
#Linear Regression
```

```
#regression model for predicting hospice days last month given % admitted last month
```

```
#as the variable
```

```
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`hospice days last month`~`% admitted last month`,  
                        data=DAPCancerEvents)
```

```
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)
```

```
#summary(DAPCancerEvents.lm)
```

```
paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)
```

```
## [1] "Adjusted r squared= 0.3401133870442"
```

predicting ICU days last month given % admitted ICU last month

```
#Linear Regression
```

```
#regression model for predicting ICU days last month considering % admitted ICU last month
```

```
#as the variable
```

```
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`ICU days last month`~DAPCancerEvents$`% admitted ICU last month`,  
                        data=DAPCancerEvents)
```

```
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)
```

```
#summary(DAPCancerEvents.lm)
```

```
paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)
```

```
## [1] "Adjusted r squared= 0.832034215365293"
```

predicting hospice days last month given % hospice last month

```
#Linear Regression
```

```
#regression model for predicting hospice days last month given percentage in hospice last month
```

#as the variable

```
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$hospice days last month~DAPCancerEvents$`% hospice last month`,
                        data=DAPCancerEvents)
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)
#summary(DAPCancerEvents.lm)
```

```
paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)
```

```
## [1] "Adjusted r squared= 0.680732429100812"
```

predicting % hospice last 3 days given % hospice last month

#Linear Regression

*#regression model for predicting % hospice last 3 days considering % hospice last month
#as the variable*

```
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`% hospice last 3 days`~DAPCancerEvents$`% hospice last month`,
                        data=DAPCancerEvents)
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)
#summary(DAPCancerEvents.lm)
```

```
paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)
```

```
## [1] "Adjusted r squared= 0.273557274772961"
```

predicting % died in hospital given hospital, hospice days and percentage hospice days last month

#Linear Regression

*#regression model for predicting % died in hospital considered both Hospital days last month and hospice days last month and % hospice last month
#as the variable*

```
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`% died in hospital`~DAPCancerEvents$`Hospital days last month`+
                        DAPCancerEvents$`% hospice last month`+DAPCancerEvents$`hospice days last month`,
                        data=DAPCancerEvents)
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)
#summary(DAPCancerEvents.lm)
```

```
paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)
```

```
## [1] "Adjusted r squared= 0.768643570928753"
```

predicting % died in hospital given state

#Linear Regression

*#regression model for predicting % died in hospital considered state
#as the variable*

```
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`% died in hospital`~DAPCancerEvents$state,
                        data=DAPCancerEvents)
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)
#summary(DAPCancerEvents.lm)
```

```
paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)
```

```
## [1] "Adjusted r squared= 0.310262953108519"
```

predicting % died in hospital given city

#Linear Regression

*#regression model for predicting % died in hospital considering city
#as the variable*

```
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`% died in  
hospital`~DAPCancerEvents$City,  
                        data=DAPCancerEvents)  
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)  
#summary(DAPCancerEvents.lm)
```

```
paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)
```

```
## [1] "Adjusted r squared= 0.439734378478089"
```

predicting % died in hospital given city and state

#Linear Regression

*#regression model for predicting % died in hospital considering city and state
#as the variable*

```
DAPCancerEvents.lm = lm(formula=DAPCancerEvents$`% died in  
hospital`~DAPCancerEvents$City+DAPCancerEvents$state,  
                        data=DAPCancerEvents)  
DAPCancerEvents.model <- summary(DAPCancerEvents.lm)  
#summary(DAPCancerEvents.lm)
```

```
paste ("Adjusted r squared=", DAPCancerEvents.model$adj.r.squared)
```

```
## [1] "Adjusted r squared= 0.518439483147694"
```

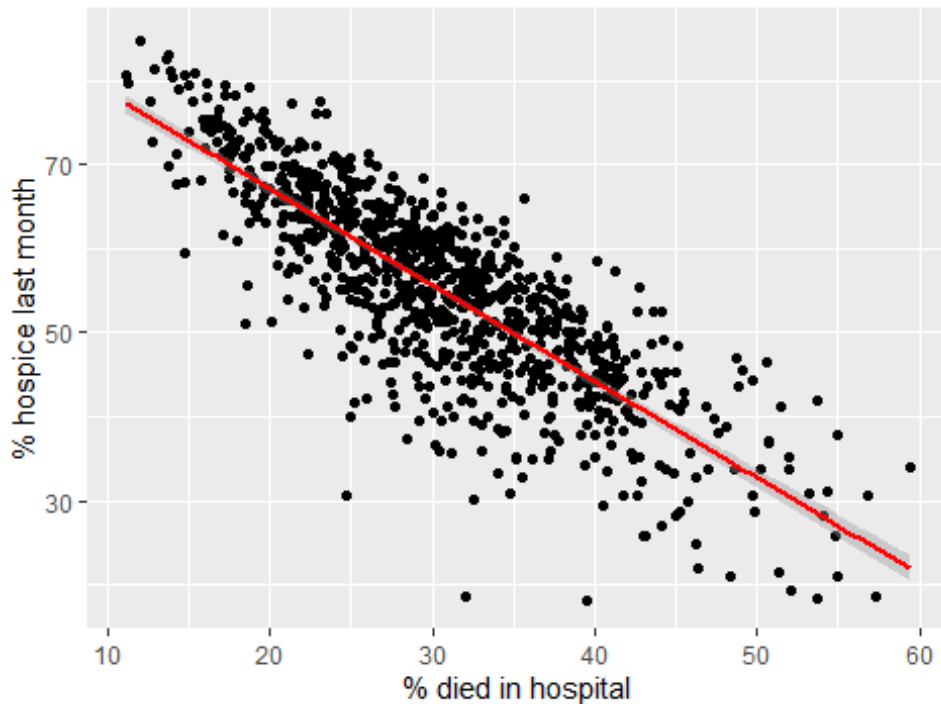
Linear Models

Column

Plot LM 1

```
g <- ggplot(DAPCancerEvents,  
            aes(x= DAPCancerEvents$`% died in hospital`,y= DAPCancerEvents$`% hospice  
last month`))  
g <- g + geom_point()  
g <- g + stat_smooth(method="lm", col = "red")  
g <- g + ggtitle("Linear Model 1: PREDICT % hospice last month BASED ON % died in  
hospital")  
g <- g+labs(x="% died in hospital", y="% hospice last month")  
g  
  
## Warning: Removed 8 rows containing non-finite values (stat_smooth).  
  
## Warning: Removed 8 rows containing missing values (geom_point).
```

Linear Model 1: PREDICT % hospice last month BASE

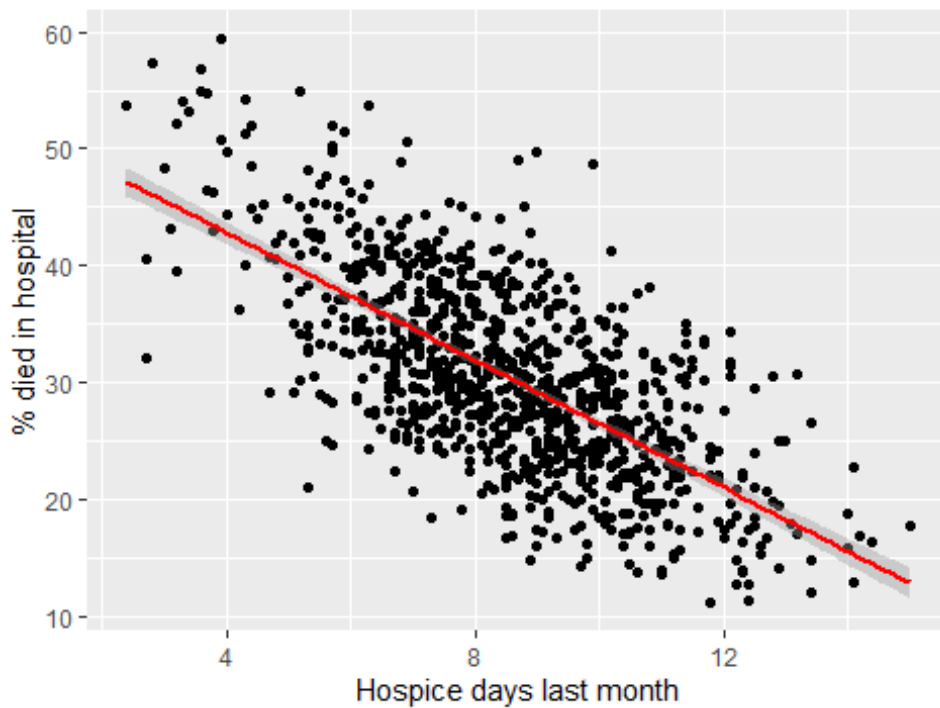


Plot LM 2

```
g <- ggplot(DAPCancerEvents, aes(x= DAPCancerEvents$`hospice days last month`,
                                y= DAPCancerEvents$`% died in hospital`)) +geom_point()
g<-g+stat_smooth(method="lm", col = "red")
g <- g + ggtitle("Linear Model 2: PREDICT % died in hospital BASED ON Hospice Days Last
Month")
g <- g+labs(x="Hospice days last month", y="% died in hospital")
g

## Warning: Removed 8 rows containing non-finite values (stat_smooth).
## Warning: Removed 8 rows containing missing values (geom_point).
```

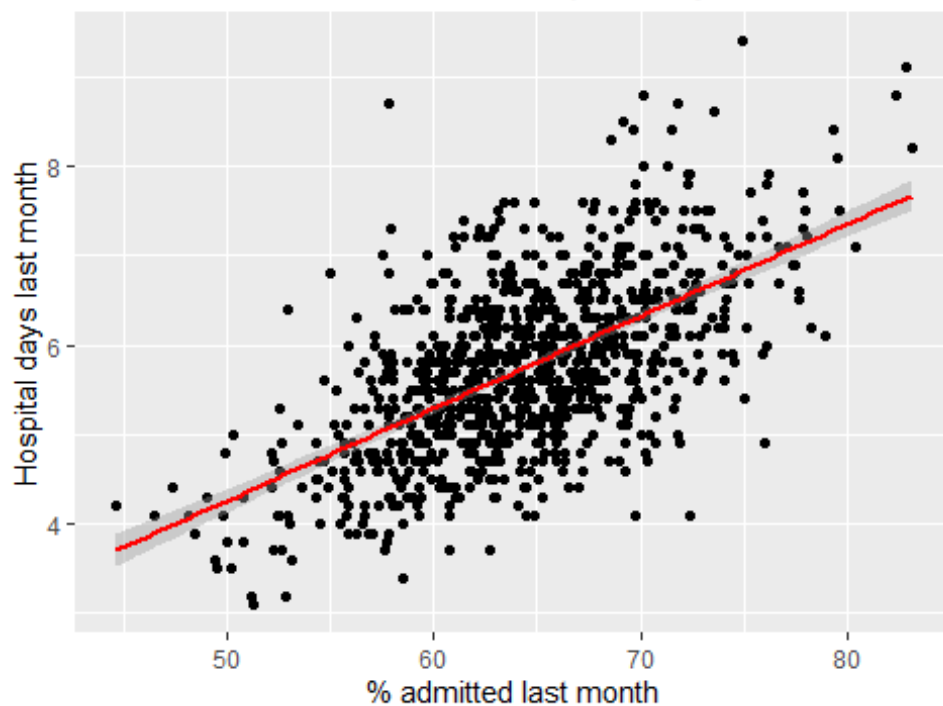
Linear Model 2: PREDICT % died in hospital BASED C



Plot LM 3

```
g <- ggplot(DAPCancerEvents, aes(x= DAPCancerEvents$`% admitted last month`,
                                y= DAPCancerEvents$`Hospital days last month`)) +
  geom_point()
g<-g+stat_smooth(method="lm", col = "red")
g<-g+ggtitle("Linear Model 3: PREDICT Hospital Days Last Month BASED ON % admitted last
month")
g <- g+labs(x="% admitted last month", y="Hospital days last month")
g
```

Linear Model 3: PREDICT Hospital Days Last Month B/

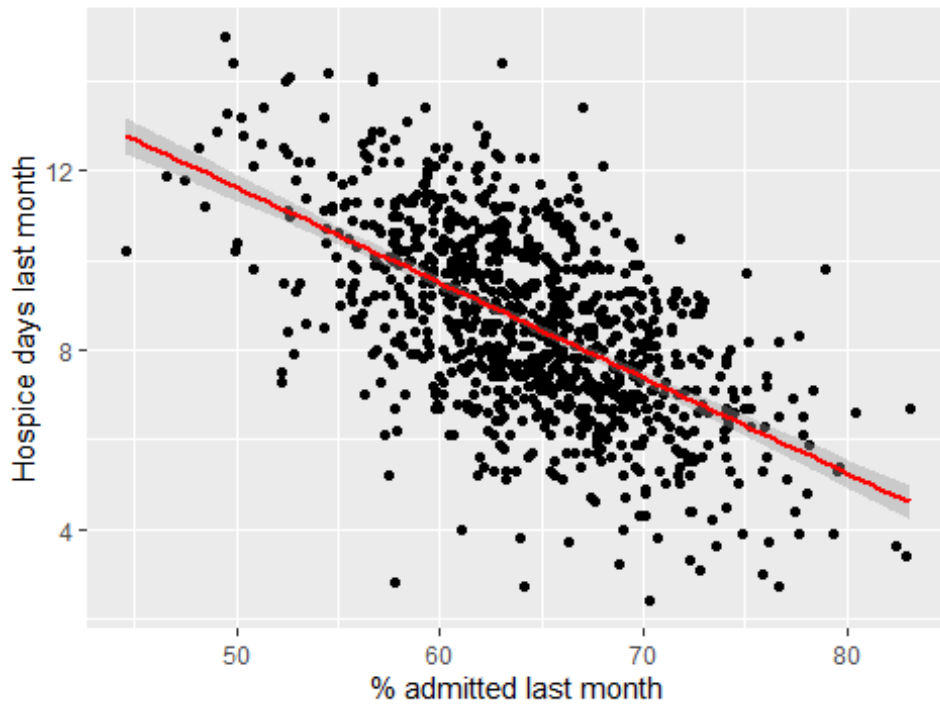


Plot LM 4

```
g<-ggplot(DAPCancerEvents,aes(x= DAPCancerEvents$`% admitted last month`, y=
DAPCancerEvents$`hospice days last month`))+geom_point()
g<-g+stat_smooth(method="lm", col = "red")
g<-g+ggtitle("Linear Model 4: PREDICT Hospice Days Last Month based on % admitted last
month")
g <- g+labs(x="% admitted last month", y="Hospice days last month")
g

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

Linear Model 4: PREDICT Hospice Days Last Month b

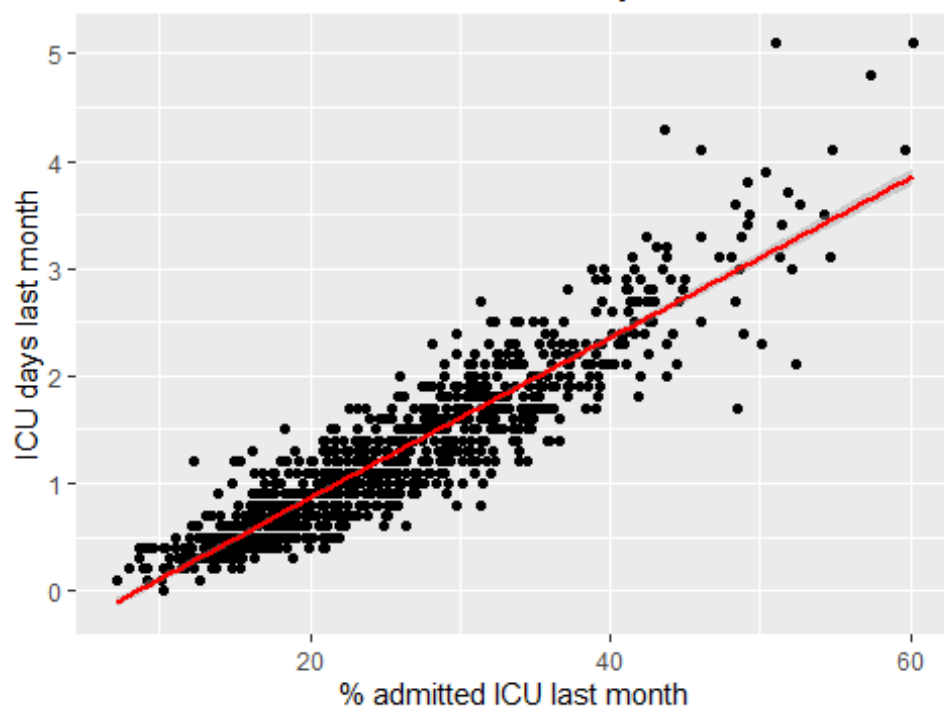


Plot LM 5

```
g<-ggplot(DAPCancerEvents,aes(x= DAPCancerEvents$`% admitted ICU last month`, y=
DAPCancerEvents$`ICU days last month`))+geom_point()
g<-g+stat_smooth(method="lm", col = "red")
g<-g+ggtitle("Linear Model 5: PREDICT ICU days last month BASED ON % admitted ICU last
month.")
g<-g+labs(x="% admitted ICU last month", y="ICU days last month")
g

## Warning: Removed 22 rows containing non-finite values (stat_smooth).
## Warning: Removed 22 rows containing missing values (geom_point).
```

Linear Model 5: PREDICT ICU days last month BASED

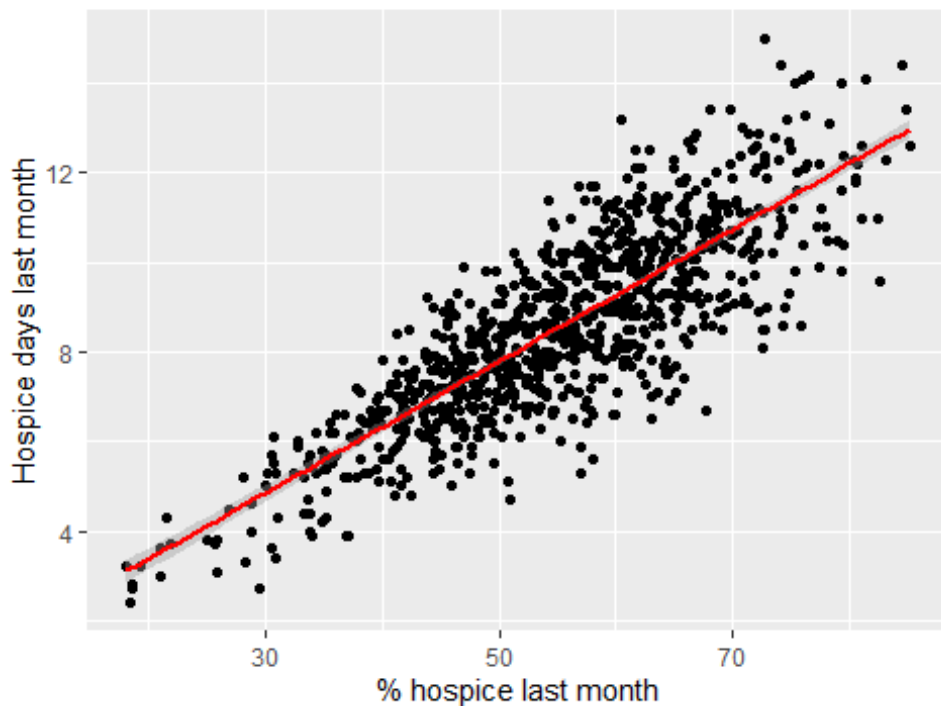


Plot LM 6

```
g<-ggplot(DAPCancerEvents,aes(x= DAPCancerEvents$`% hospice last month`, y=
DAPCancerEvents$`hospice days last month`))+geom_point()
g<-g+stat_smooth(method="lm", col = "red")
g<-g+ggtitle("Linear Model 6: PREDICT hospice days last month based on % hospice last
month")
g <- g+labs(x="% hospice last month", y="Hospice days last month")
g

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

Linear Model 6: PREDICT hospice days last month bas



Plot LM 7

```
g<-ggplot(DAPCancerEvents,aes(x= DAPCancerEvents$`% hospice last month`, y=
DAPCancerEvents$`% hospice last 3 days`))+geom_point()
g<-g+stat_smooth(method="lm", col = "red")
g<-g+ggtitle("Linear Model 7: PREDICT hospice last 3 days of life based on % hospice last
month")
g <- g+labs(x="% hospice last 3 days", y="% hospice last month")
g

## Warning: Removed 446 rows containing non-finite values (stat_smooth).
## Warning: Removed 446 rows containing missing values (geom_point).
```

Linear Model 7: PREDICT hospice last 3 days of life ba

