DSC680 Project2 - Vet Hospital Wait Times - Part1

Amie Davis

23 April, 2021

Data Sources:

Dove Lewis Animal Hospital, Portland, OR, 1Jan2019-11Apr2021, Proprietary Data

Load Libraries

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.0.2
```

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.0.2
```

```
library(ggplot2)
library(lubridate)
library(dplyr)
#require(scales)
```

1. Prepare Data

a) Import the Data

```
# Column range required for all sheets

# Load Whiteboard data
board_df <- read_excel("Data/WhiteBoard Tracker.xlsx", range = cell_cols("A:K"))

# Load Wait data
# Not loading correctly with read_excel, so switched to read_csv and reformatted dates
wait_df <- read_csv("Data/Smart Flow Wait Times.csv",
                    col_types = cols_only(
#                                           "Clinical ID" = col_double(),
                                           "Time Stamp" = col_datetime(format="%m/%d/%y %H:%M"),
                                           "Patient Count" = col_integer(),
                                           "Wait Time Average" = col_double(),
                                           "Time Block" = col_integer(),
                                           "Window" = col_character()
                                           ))

# Load Patient data
# Converted Excel file to csv using Trifacta to split data fields that were delimited
patient_df <- read_csv("Data/Patient Snapshot_Revised.csv",
                         col_types = cols_only("Department" = col_character(),
                                           "Consult Date" = col_date(format="%m/%d/%Y"),
                                           "Consult Division" = col_character(),
                                           "Clinical Number" = col_integer(),
                                           "Patient Number" = col_integer(),
                                           "Triage Type" = col_integer(),
                                           "Clinical Description" = col_character(),
                                           "Appointment Type1" = col_character(),
                                           "Appointment Type2" = col_character(),
                                           "Appointment Type3" = col_character(),
                                           "Appointment Date1" = col_datetime(format="%m/%d/%Y %H:%M"),
                                           "Appointment Date2" = col_datetime(format="%m/%d/%Y %H:%M"),
                                           "Appointment Date3" = col_datetime(format="%m/%d/%Y %H:%M"),
                                           "Presenting Problem1" = col_character(),
                                           "Presenting Problem2" = col_character(),
                                           "Presenting Problem3" = col_character(),
                                           "Therapeutic-Procedure1" = col_character(),
                                           "Therapeutic-Procedure2" = col_character(),
                                           "Therapeutic-Procedure3" = col_character(),
                                           "Therapeutic-Procedure4" = col_character(),
                                           "Therapeutic-Procedure5" = col_character(),
                                           "Therapeutic-Procedure6" = col_character(),
                                           "Therapeutic-Procedure7" = col_character(),
                                           "Therapeutic-Procedure8" = col_character()
                                           ))
```

b) Review Features

```
str(board_df)
```

```
str(wait_df)
```

```
str(patient_df)
```

c) Derived Features

```
# Convert date fields to date stamps
# Split date fields into separate columns using lubridate package
# Time Stamp is POSIXlt, so need to convert to date first
board_df$Date_Stamp <- date(board_df$"Time Stamp")
board_df$TS_HOUR <- hour(board_df$"Time Stamp")
board_df2 <- board_df %>% mutate(Date_Stamp = ymd(Date_Stamp))
board_df_new <- board_df2 %>% mutate (TS_YEAR = year(Date_Stamp),
                                      TS_MONTH = month(Date_Stamp),
                                      TS_DAY = day(Date_Stamp),
                                      TS_DOW = weekdays(Date_Stamp),
                                      TS_WEEK = week(Date_Stamp)
                                      )

wait_df$Date_Stamp <- date(wait_df$"Time Stamp")
wait_df$TS_HOUR <- hour(wait_df$"Time Stamp")
```

```
wait_df2 <- wait_df %>% mutate(Date_Stamp = ymd(Date_Stamp))
wait_df_new <- wait_df2 %>% mutate (TS_YEAR = year(Date_Stamp),
                                     TS_MONTH = month(Date_Stamp),
                                     TS_DAY = day(Date_Stamp),
                                     TS_DOW = weekdays(Date_Stamp),
                                     TS_WEEK = week(Date_Stamp)
                                     )

# Convert categorical variables to factors
board_df_new$TS_YEAR <- factor(board_df_new$TS_YEAR)
board_df_new$TS_MONTH <- factor(board_df_new$TS_MONTH)
board_df_new$TS_DAY <- factor(board_df_new$TS_DAY)
board_df_new$TS_HOUR <- factor(board_df_new$TS_HOUR)
board_df_new$TS_DOW <- factor(board_df_new$TS_DOW)

wait_df_new$TS_YEAR <- factor(wait_df_new$TS_YEAR)
wait_df_new$TS_MONTH <- factor(wait_df_new$TS_MONTH)
wait_df_new$TS_DAY <- factor(wait_df_new$TS_DAY)
wait_df_new$TS_HOUR <- factor(wait_df_new$TS_HOUR)
wait_df_new$TS_DOW <- factor(wait_df_new$TS_DOW)
wait_df_new$"Time Block" <- factor(wait_df_new$"Time Block")
wait_df_new$Window <- factor(wait_df_new$Window)

patient_df$Department <- factor(patient_df$Department)
patient_df$"Consult Division" <- factor(patient_df$"Consult Division")
patient_df$"Triage Type" <- factor(patient_df$"Triage Type")
patient_df$"Therapeutic-Procedure1" <- factor(patient_df$"Therapeutic-Procedure1")
patient_df$"Therapeutic-Procedure2" <- factor(patient_df$"Therapeutic-Procedure2")
patient_df$"Therapeutic-Procedure3" <- factor(patient_df$"Therapeutic-Procedure3")
patient_df$"Therapeutic-Procedure4" <- factor(patient_df$"Therapeutic-Procedure4")
patient_df$"Therapeutic-Procedure5" <- factor(patient_df$"Therapeutic-Procedure5")
patient_df$"Therapeutic-Procedure6" <- factor(patient_df$"Therapeutic-Procedure6")
patient_df$"Therapeutic-Procedure7" <- factor(patient_df$"Therapeutic-Procedure7")
patient_df$"Therapeutic-Procedure8" <- factor(patient_df$"Therapeutic-Procedure8")
patient_df$"Appointment Type1" <- factor(patient_df$"Appointment Type1")
patient_df$"Appointment Type2" <- factor(patient_df$"Appointment Type2")
patient_df$"Appointment Type3" <- factor(patient_df$"Appointment Type3")
patient_df$"Presenting Problem1" <- factor(patient_df$"Presenting Problem1")
patient_df$"Presenting Problem2" <- factor(patient_df$"Presenting Problem2")
patient_df$"Presenting Problem3" <- factor(patient_df$"Presenting Problem3")


head(board_df_new)
```

```
## # A tibble: 6 x 18
##   `Row ID` `Outpatient Cou~ `ICU Patient Co~ `Time Stamp`
##      <dbl>            <dbl>            <dbl> <dttm>
## 1   254608               11               11 2019-01-01 00:08:38
## 2   254609               11               11 2019-01-01 00:13:45
## 3   254610               11               11 2019-01-01 00:18:53
## 4   254611                9               11 2019-01-01 00:24:03
## 5   254612                9               11 2019-01-01 00:29:11
## 6   254613                9               11 2019-01-01 00:34:13
## # ... with 14 more variables: Time <dttm>, `TIME Hour` <dttm>, Weekday <dttm>,
## #   Date <dttm>, Year <dbl>, Week <dbl>, Month <dbl>, Date_Stamp <date>,
## #   TS_HOUR <fct>, TS_YEAR <fct>, TS_MONTH <fct>, TS_DAY <fct>, TS_DOW <fct>,
## #   TS_WEEK <dbl>
```

```
head(wait_df_new)
```

```
## # A tibble: 6 x 12
##   `Time Stamp`        `Patient Count` `Wait Time Aver~ `Time Block` Window
##   <dttm>                        <int>            <dbl> <fct>        <fct>
## 1 2020-08-10 00:04:00               9              188 1            12am--~
## 2 2020-08-10 00:09:00               9              193 1            12am--~
## 3 2020-08-10 00:14:00               9              197 1            12am--~
## 4 2020-08-10 00:19:00               9              202 1            12am--~
## 5 2020-08-10 00:24:00              10              188 1            12am--~
## 6 2020-08-10 00:29:00              10              192 1            12am--~
## # ... with 7 more variables: Date_Stamp <date>, TS_HOUR <fct>, TS_YEAR <fct>,
## #   TS_MONTH <fct>, TS_DAY <fct>, TS_DOW <fct>, TS_WEEK <dbl>
```

```
head(patient_df)
```

```
## # A tibble: 6 x 24
##   Department `Consult Date` `Consult Divisi~ `Clinical Numbe~ `Patient Number`
##   <fct>      <date>         <fct>                       <int>            <int>
## 1 A-ECC      2018-07-02     ECC                        901219           298033
## 2 A-ECC      2018-07-02     ECC                        901218           298032
## 3 A-ECC      2018-07-02     ECC                        901220           298034
## 4 A-ECC      2018-07-02     ECC                        901221           287042
## 5 A-ECC      2018-07-02     ECC                        901222           298035
## 6 A-ECC      2018-07-02     ECC                        901223           298036
## # ... with 19 more variables: `Triage Type` <fct>, `Clinical
## #   Description` <chr>, `Appointment Type1` <fct>, `Appointment Type2` <fct>,
## #   `Appointment Type3` <fct>, `Appointment Date1` <dttm>, `Appointment
## #   Date2` <dttm>, `Appointment Date3` <dttm>, `Presenting Problem1` <fct>,
## #   `Presenting Problem2` <fct>, `Presenting Problem3` <fct>,
## #   `Therapeutic-Procedure1` <fct>, `Therapeutic-Procedure2` <fct>,
## #   `Therapeutic-Procedure3` <fct>, `Therapeutic-Procedure4` <fct>,
## #   `Therapeutic-Procedure5` <fct>, `Therapeutic-Procedure6` <fct>,
## #   `Therapeutic-Procedure7` <fct>, `Therapeutic-Procedure8` <fct>
```

d) Summary Statistics

```
#str(board_df_new)
summary(board_df_new)
```

Observations:

181,742 Whiteboard Records

No N/A values or outliers

```
#str(wait_df_new)
summary(wait_df_new)
```

Observations:

64,032 Smart Flow Wait Records

No N/A values

Extreme estimated wait times confirmed by source

```
str(patient_df)
#summary(patient_df)
```

Observations:

74,529 Weekly Client Records

68,291 Records do not have a Triage Type.

- Hospital did not start tracking Triages until 4 Jan 2021

- Triages are tracked for emergency room only.

26 Records do not have an Appointment Type.

89 Records do not have an Appointment Date.

74,422 Records do not have a Presenting Problem recorded.

56,511 Records do not have a Procedure listed.

e) Drop unneeded columns

```
# Drop unused pre-calculated datetime fields

board_df_new[ ,c(
  "Time",
  "TIME Hour",
  "Weekday",
  "Date",
  "Year",
  "Week",
  "Month"
)] <- list(NULL)
```

f) Remove Outliers

```
# If no 1st appt date, use 2nd appt date
# Note that ifelse changes type, so using dplr's if_else
patient_df$Appt_Date <- if_else(is.na(patient_df$"Appointment Date1"), patient_df$"Appointment Date2", p

# Only use patient records w appt date
clean_data<-subset(patient_df,!(is.na(patient_df["Appt_Date"])))

# Only use patient records w appt type
```

```
clean_data<-subset(clean_data,!(is.na(clean_data["Appointment Type1"])))

# Look specific record number 901362
#test_data<-subset(clean_data,clean_data$"Clinical Number" == 901362)
#head(test_data)

#summary(clean_data)
#str(clean_data)
#74,479 records remaining
```

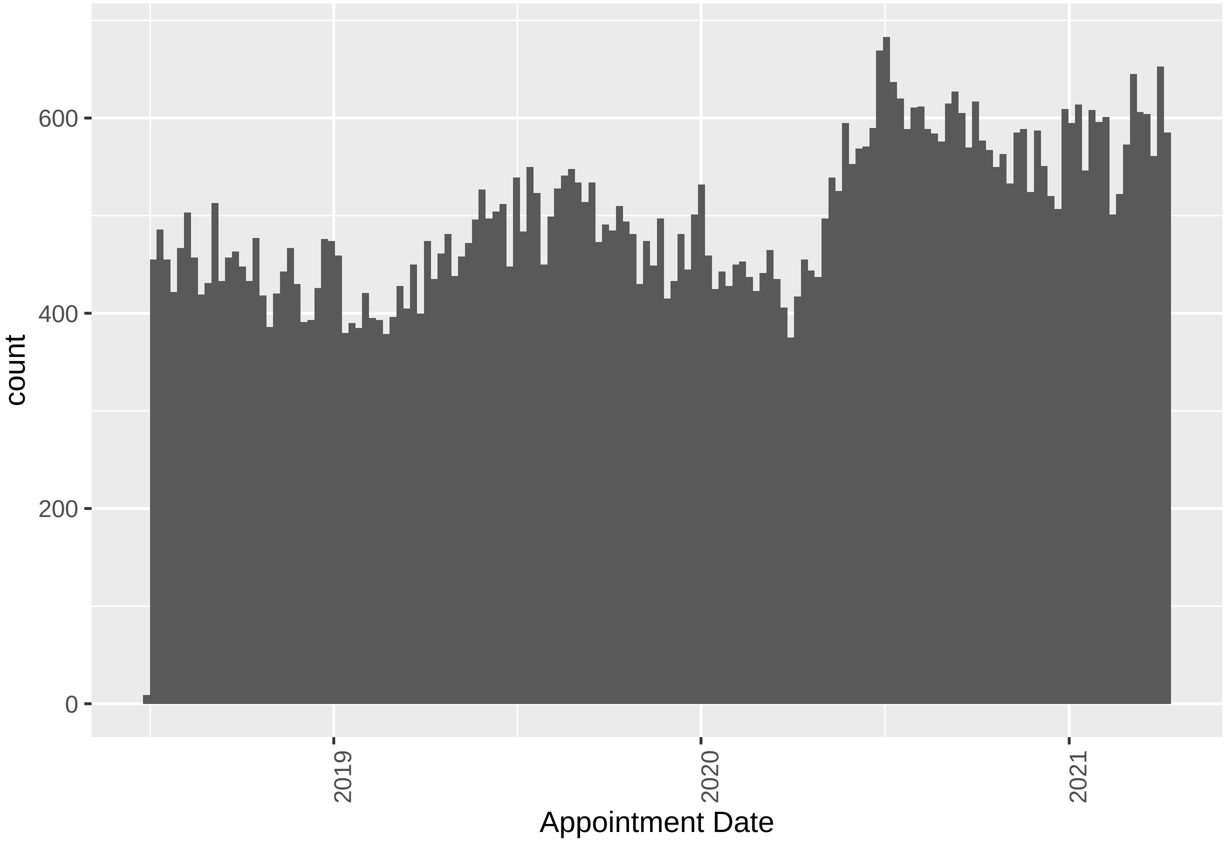
g) Export dataset

```
# Export data for use in visualizations
write.csv(clean_data,"Data/patient_new.csv")
```

2. EDA - Review Distributions

a) Plot Histograms for Numeric Vars

```
ggplot(clean_data, aes(x=Appt_Date)) +
    geom_histogram(bins=150) +
    labs(x="Appointment Date") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

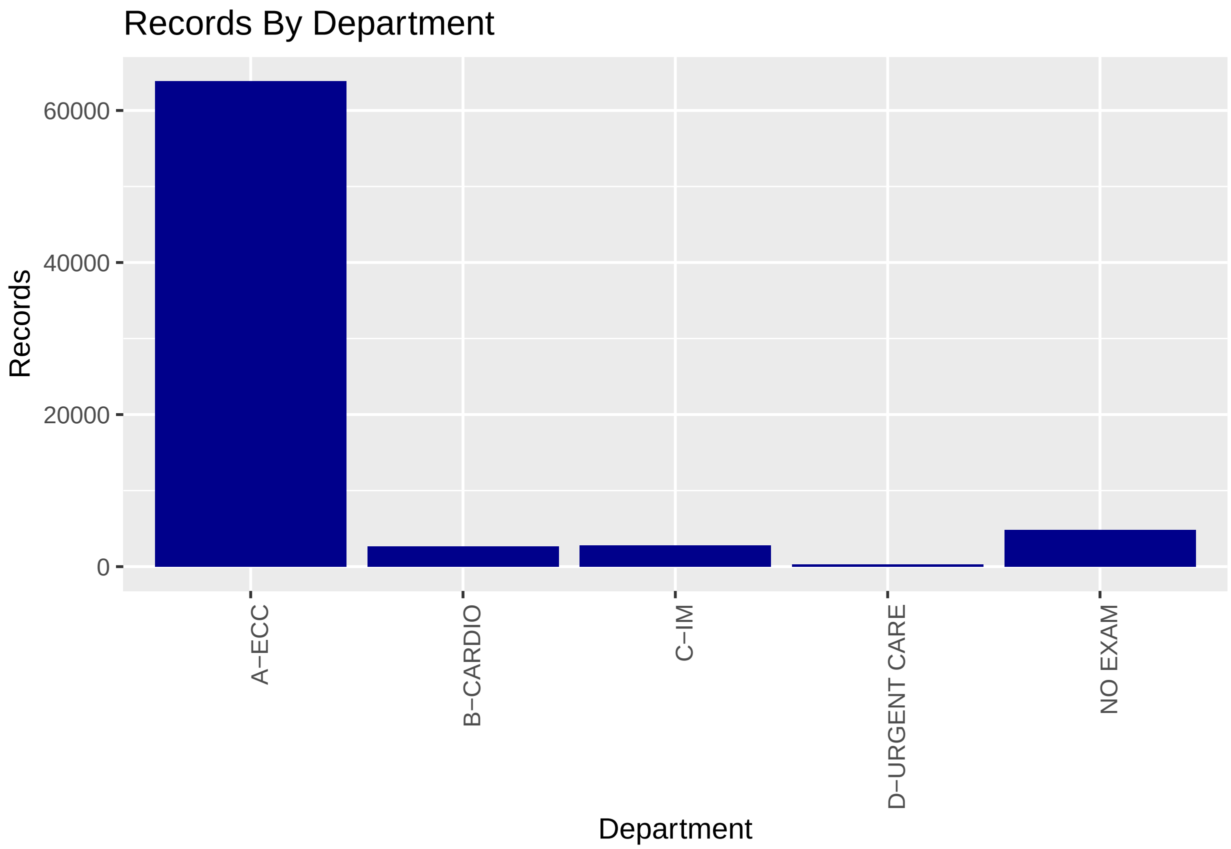


There appears to be a peak season every year.

b) Histograms for Categorical Features

```
# Department
p <- ggplot(clean_data, aes(x=Department)) +
     geom_bar(fill="dark blue") +
     labs(x="Department", y="Records", title="Records By Department") +
     theme(axis.text.x = element_text(angle = 90, hjust = 1))

p
```



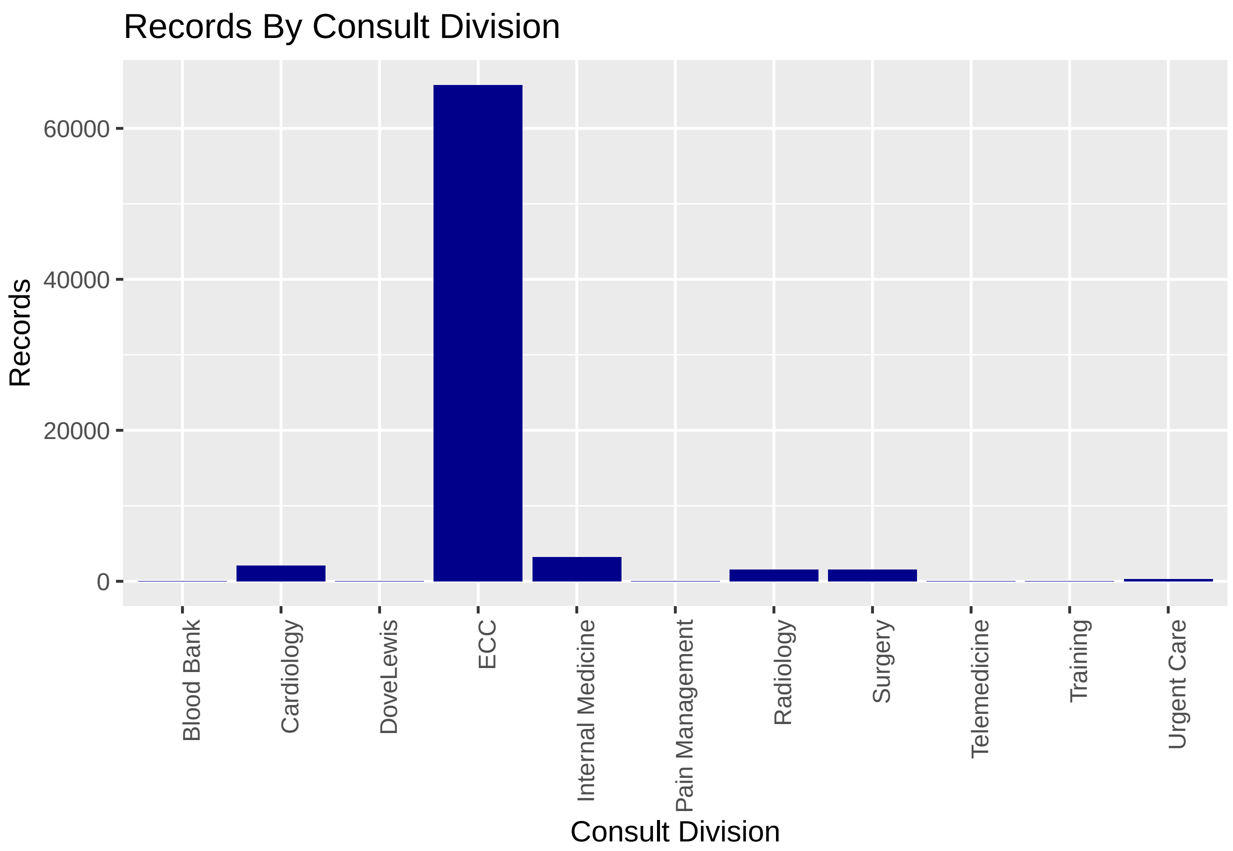
As you can see, most appointments are in the Emergency Critical Care Department. This includes both Outpatient and ICU patients.

```
# Offense Codes

p <- ggplot(clean_data, aes(x=clean_data$"Consult Division")) +
     geom_bar(fill="dark blue") +
     labs(x="Consult Division", y="Records", title="Records By Consult Division") +
     theme(axis.text.x = element_text(angle = 90, hjust = 1))

p
```

```
## Warning: Use of `clean_data$"Consult Division"` is discouraged. Use `Consult
## Division` instead.
```



This is expected. Most appointments are in the Emergency Critical Care Division.