library(tidyverse)

## -- Attaching packages ----------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.0 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.5  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts -------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

Telecom <- read\_csv("Telecom.csv")

## Parsed with column specification:  
## cols(  
## customerID = col\_character(),  
## MonthlyCharges = col\_double(),  
## TotalCharges = col\_character(),  
## PaymentMethod = col\_character(),  
## Churn = col\_character()  
## )

Telecom$MonthlyCharges

## [1] 29.85 56.95 NA 42.30 70.70 NaN 89.10 NA 104.80 54.10

is.na(Telecom$MonthlyCharges)

## [1] FALSE FALSE TRUE FALSE FALSE TRUE FALSE TRUE FALSE FALSE

summarize(Telecom, count = sum(is.na(MonthlyCharges)))

## # A tibble: 1 x 1  
## count  
## <int>  
## 1 3

mutate(Telecom, MonthlyCharges = replace(MonthlyCharges, is.na(MonthlyCharges),  
 median(MonthlyCharges, na.rm = TRUE)))

## # A tibble: 10 x 5  
## customerID MonthlyCharges TotalCharges PaymentMethod Churn  
## <chr> <dbl> <chr> <chr> <chr>  
## 1 7590-VHVEG 29.8 109.9 Electronic check yes   
## 2 5575-GNVDE 57.0 na Mailed check yes   
## 3 3668-QPYBK 57.0 108.15 -- yes   
## 4 7795-CFOCW 42.3 1840.75 Bank transfer no   
## 5 9237-HQITU 70.7 <NA> Electronic check no   
## 6 9305-CDSKC 57.0 820.5 -- yes   
## 7 1452-KIOVK 89.1 1949.4 Credit card no   
## 8 6713-OKOMC 57.0 N/A <NA> yes   
## 9 7892-POOKP 105. 3046.05 Electronic check no   
## 10 8451-AJOMK 54.1 354.95 Electronic check no

Telecom

## # A tibble: 10 x 5  
## customerID MonthlyCharges TotalCharges PaymentMethod Churn  
## <chr> <dbl> <chr> <chr> <chr>  
## 1 7590-VHVEG 29.8 109.9 Electronic check yes   
## 2 5575-GNVDE 57.0 na Mailed check yes   
## 3 3668-QPYBK NA 108.15 -- yes   
## 4 7795-CFOCW 42.3 1840.75 Bank transfer no   
## 5 9237-HQITU 70.7 <NA> Electronic check no   
## 6 9305-CDSKC NaN 820.5 -- yes   
## 7 1452-KIOVK 89.1 1949.4 Credit card no   
## 8 6713-OKOMC NA N/A <NA> yes   
## 9 7892-POOKP 105. 3046.05 Electronic check no   
## 10 8451-AJOMK 54.1 354.95 Electronic check no

Telecom <- mutate(Telecom, MonthlyCharges = replace(MonthlyCharges, is.na(MonthlyCharges),median(MonthlyCharges, na.rm = TRUE)))  
  
Telecom

## # A tibble: 10 x 5  
## customerID MonthlyCharges TotalCharges PaymentMethod Churn  
## <chr> <dbl> <chr> <chr> <chr>  
## 1 7590-VHVEG 29.8 109.9 Electronic check yes   
## 2 5575-GNVDE 57.0 na Mailed check yes   
## 3 3668-QPYBK 57.0 108.15 -- yes   
## 4 7795-CFOCW 42.3 1840.75 Bank transfer no   
## 5 9237-HQITU 70.7 <NA> Electronic check no   
## 6 9305-CDSKC 57.0 820.5 -- yes   
## 7 1452-KIOVK 89.1 1949.4 Credit card no   
## 8 6713-OKOMC 57.0 N/A <NA> yes   
## 9 7892-POOKP 105. 3046.05 Electronic check no   
## 10 8451-AJOMK 54.1 354.95 Electronic check no

# replacing with standard missing value type, NA  
Telecom <- mutate(Telecom, TotalCharges = replace(TotalCharges, TotalCharges == "na", NA))  
  
Telecom <- mutate(Telecom, TotalCharges = replace(TotalCharges, TotalCharges == "N/A", NA))  
  
#notice that this particular column came over as chr. Lets change to numeric.  
Telecom$TotalCharges <- as.numeric(Telecom$TotalCharges)  
Telecom

## # A tibble: 10 x 5  
## customerID MonthlyCharges TotalCharges PaymentMethod Churn  
## <chr> <dbl> <dbl> <chr> <chr>  
## 1 7590-VHVEG 29.8 110. Electronic check yes   
## 2 5575-GNVDE 57.0 NA Mailed check yes   
## 3 3668-QPYBK 57.0 108. -- yes   
## 4 7795-CFOCW 42.3 1841. Bank transfer no   
## 5 9237-HQITU 70.7 NA Electronic check no   
## 6 9305-CDSKC 57.0 820. -- yes   
## 7 1452-KIOVK 89.1 1949. Credit card no   
## 8 6713-OKOMC 57.0 NA <NA> yes   
## 9 7892-POOKP 105. 3046. Electronic check no   
## 10 8451-AJOMK 54.1 355. Electronic check no

Telecom <- mutate(Telecom, TotalCharges = replace(TotalCharges, is.na(TotalCharges),mean(TotalCharges, na.rm = TRUE)))  
  
Telecom

## # A tibble: 10 x 5  
## customerID MonthlyCharges TotalCharges PaymentMethod Churn  
## <chr> <dbl> <dbl> <chr> <chr>  
## 1 7590-VHVEG 29.8 110. Electronic check yes   
## 2 5575-GNVDE 57.0 1176. Mailed check yes   
## 3 3668-QPYBK 57.0 108. -- yes   
## 4 7795-CFOCW 42.3 1841. Bank transfer no   
## 5 9237-HQITU 70.7 1176. Electronic check no   
## 6 9305-CDSKC 57.0 820. -- yes   
## 7 1452-KIOVK 89.1 1949. Credit card no   
## 8 6713-OKOMC 57.0 1176. <NA> yes   
## 9 7892-POOKP 105. 3046. Electronic check no   
## 10 8451-AJOMK 54.1 355. Electronic check no

Telecom <- mutate(Telecom, PaymentMethod = replace(PaymentMethod, PaymentMethod == "--", NA))  
  
Telecom <- mutate(Telecom,PaymentMethod = replace(PaymentMethod, is.na(PaymentMethod), "unavailable"))  
  
Telecom

## # A tibble: 10 x 5  
## customerID MonthlyCharges TotalCharges PaymentMethod Churn  
## <chr> <dbl> <dbl> <chr> <chr>  
## 1 7590-VHVEG 29.8 110. Electronic check yes   
## 2 5575-GNVDE 57.0 1176. Mailed check yes   
## 3 3668-QPYBK 57.0 108. unavailable yes   
## 4 7795-CFOCW 42.3 1841. Bank transfer no   
## 5 9237-HQITU 70.7 1176. Electronic check no   
## 6 9305-CDSKC 57.0 820. unavailable yes   
## 7 1452-KIOVK 89.1 1949. Credit card no   
## 8 6713-OKOMC 57.0 1176. unavailable yes   
## 9 7892-POOKP 105. 3046. Electronic check no   
## 10 8451-AJOMK 54.1 355. Electronic check no