# Imputation on Time Series Missing Data - MICE

IDS.506: Fargo Health Group Assignment Technical Appendix Rmd Notebook 2

Robert Duc Bui - mbui7@uic.edu - 660809303

```
library(tidyverse)
library(tidymodels)
library(lubridate)
library(Cairo)
library(mice)
library(imputeTS)
```

```
raw_ts <- read_csv("data/raw_ts.csv",
                   show_col_types = FALSE) %>%
  transmute(
    y = `Incoming Examinations`,
    Year = Year,
    Month = Month
  )
```

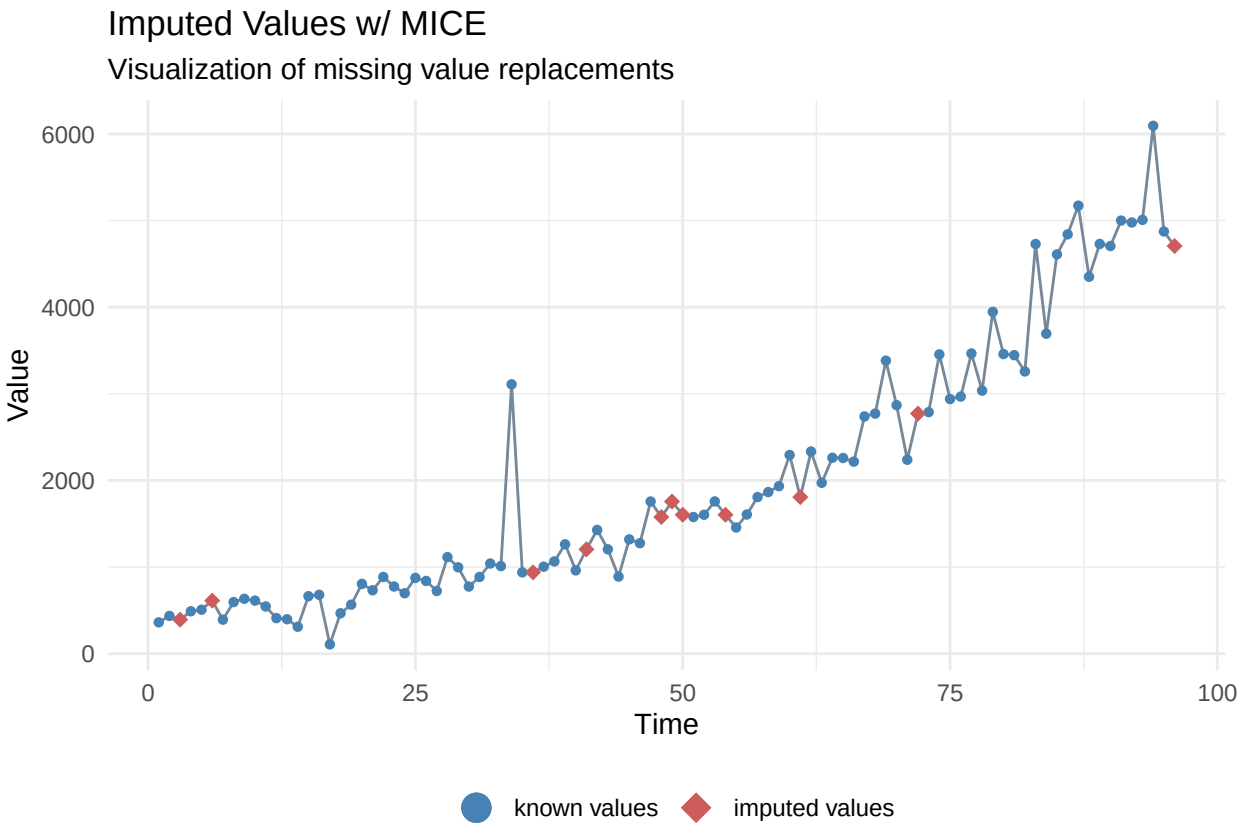
```
# MICE w/ 50 iterations of 50 imputations
obj_mice <- raw_ts %>%
  mice(m = 50,
       maxit = 50,
       seed = 8675309,
       printFlag = FALSE)

df_mice <- obj_mice %>% complete(1)

df_mice
```

```
##       y Year Month
## 1   362 2006     1
## 2   436 2006     2
## 3   393 2006     3
## 4   490 2006     4
## 5   508 2006     5
## 6   613 2006     6
## 7   393 2006     7
## 8   596 2006     8
## 9   634 2006     9
## 10  613 2006    10
## 11  545 2006    11
## 12  411 2006    12
## 13  398 2007     1
## 14  311 2007     2
```

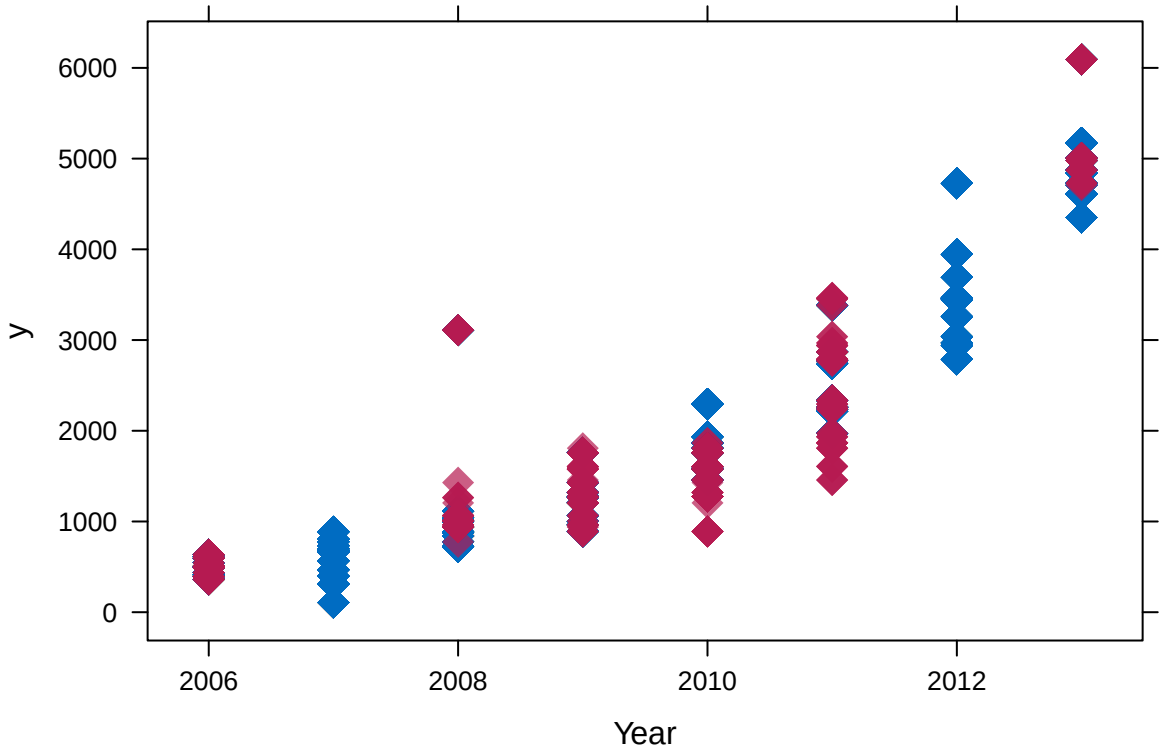
```
## 15  664 2007     3
## 16  680 2007     4
## 17  107 2007     5
## 18  467 2007     6
## 19  566 2007     7
## 20  806 2007     8
## 21  732 2007     9
## 22  886 2007    10
## 23  776 2007    11
## 24  698 2007    12
## 25  875 2008     1
## 26  840 2008     2
## 27  724 2008     3
## 28 1115 2008     4
## 29  997 2008     5
## 30  775 2008     6
## 31  886 2008     7
## 32 1041 2008     8
## 33 1011 2008     9
## 34 3110 2008    10
## 35  939 2008    11
## 36  939 2008    12
## 37 1004 2009     1
## 38 1065 2009     2
## 39 1263 2009     3
## 40  962 2009     4
## 41 1205 2009     5
## 42 1429 2009     6
## 43 1205 2009     7
## 44  890 2009     8
## 45 1320 2009     9
## 46 1276 2009    10
## 47 1757 2009    11
## 48 1578 2009    12
## 49 1757 2010     1
## 50 1604 2010     2
## 51 1578 2010     3
## 52 1604 2010     4
## 53 1758 2010     5
## 54 1604 2010     6
## 55 1457 2010     7
## 56 1607 2010     8
## 57 1808 2010     9
## 58 1866 2010    10
## 59 1934 2010    11
## 60 2294 2010    12
## 61 1808 2011     1
## 62 2334 2011     2
## 63 1973 2011     3
## 64 2262 2011     4
## 65 2259 2011     5
## 66 2217 2011     6
## 67 2739 2011     7
## 68 2772 2011     8
```

```
## 69 3383 2011     9
## 70 2869 2011    10
## 71 2239 2011    11
## 72 2772 2011    12
## 73 2789 2012     1
## 74 3455 2012     2
## 75 2940 2012     3
## 76 2968 2012     4
## 77 3466 2012     5
## 78 3037 2012     6
## 79 3946 2012     7
## 80 3459 2012     8
## 81 3446 2012     9
## 82 3258 2012    10
## 83 4729 2012    11
## 84 3694 2012    12
## 85 4610 2013     1
## 86 4841 2013     2
## 87 5172 2013     3
## 88 4351 2013     4
## 89 4730 2013     5
## 90 4706 2013     6
## 91 5000 2013     7
## 92 4978 2013     8
## 93 5008 2013     9
## 94 6094 2013    10
## 95 4874 2013    11
## 96 4706 2013    12
```

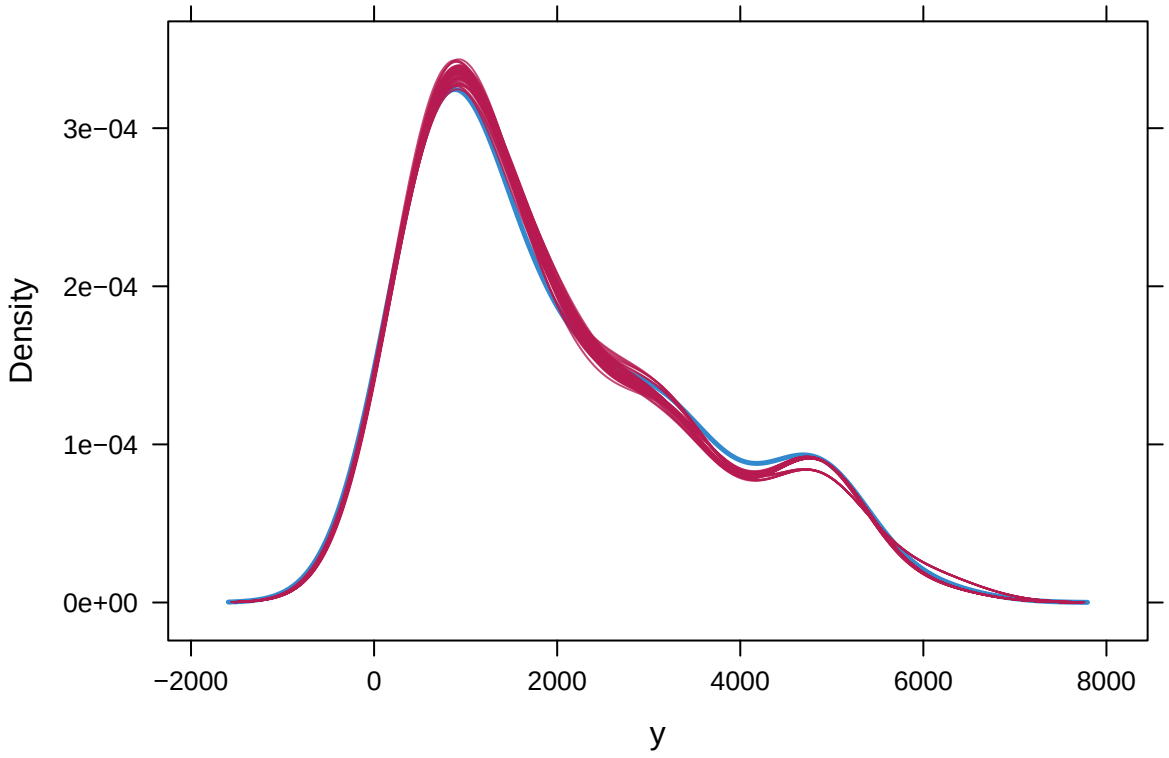
```
ts <- raw_ts %>%
  transmute(
    y = y,
    datetime = paste(Year,Month,"01",sep="-") %>% ymd()
  ) %>% select(y) %>%
  ts(start = c(2006,1),
     end   = c(2013,12),
     frequency = 12)

ts_mice <- df_mice %>%
  transmute(
    y = y,
    datetime = paste(Year,Month,"01",sep="-") %>% ymd()
  ) %>% select(y) %>%
  ts(start = c(2006,1),
     end   = c(2013,12),
     frequency = 12)

ggplot_na_imputations(x_with_na = ts,
                      x_with_imputations = ts_mice,
                      title = "Imputed Values w/ MICE",
                      theme = ggplot2::theme_minimal())
```




```
# Dotplot of imputed (red) vs real (blue) per year
xyplot(obj_mice, y ~ Year, pch = 18, cex = 2)
```




```
# Density plot of imputed data
densityplot(obj_mice, n = 96)
```




```
df_mice %>%
  rename(
    `Incoming Examinations` = y
  ) %>%
  write_csv("data/ts_mice.csv")
```