

Discontinuity Regression Samples

Austin

5/7/2019

```
library(tidyverse)
library(stringr)
library(here)
library(lubridate)
library(nlme)
library(ggrepel)

standard <- read_csv("C:/Users/Austin/Documents/Academics/NCF/Semester_2/Distributed/Proj_2/Airlines_rej

# Removes stupid government tallies
standard <- standard %>%
  filter(!str_detect(Date, 'Sub-Total'))

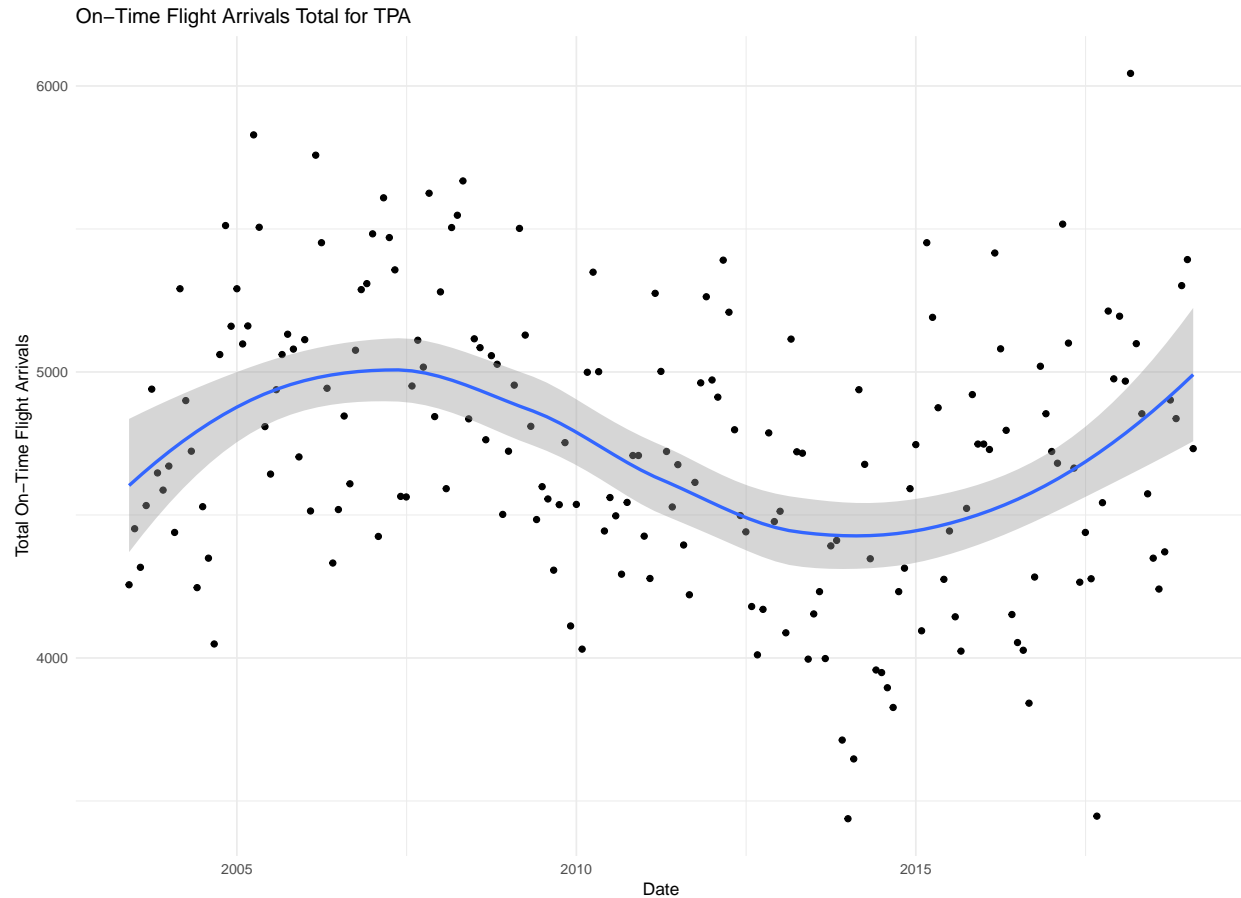
# Converts Month-Year (JUN-13) to Date Format (2013-06-01)
standard$Date <- as.Date(sub("\\\\-", "01", standard$Date), "%b%d%y")
```

Discontinuity

```
# Getting general arrival demand shape by filtering to TPA:
standard_tpa <- standard %>%
  filter(Facility == 'TPA')

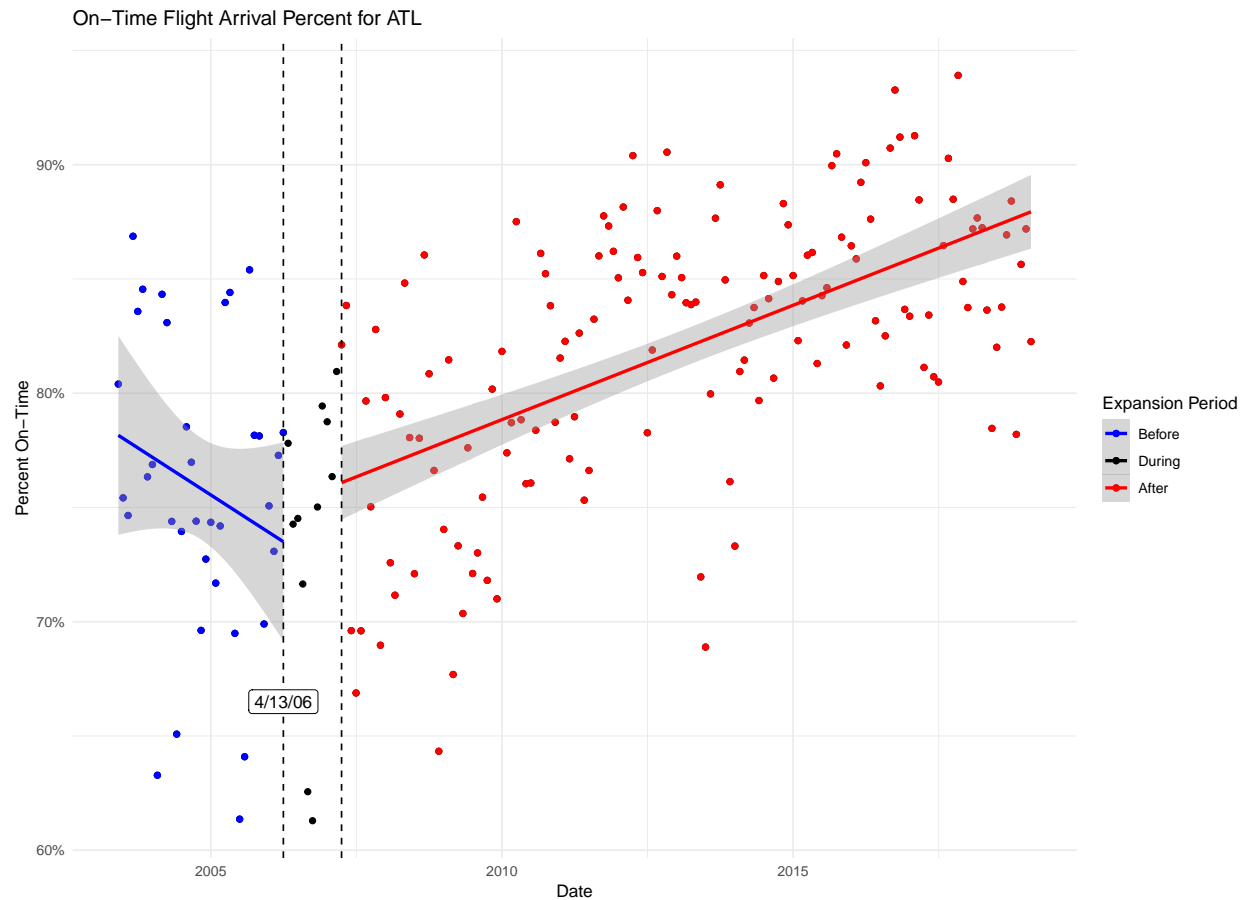
# Total On-Time Arrivals (OPS) for TPA:
ggplot(standard_tpa, aes(x=Date, y=On_Time_Arrivals)) +
  geom_point() +
  geom_smooth() +
  ggtitle(label = "On-Time Flight Arrivals Total for TPA") +
  labs(y = "Total On-Time Flight Arrivals") +
  theme_minimal()

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

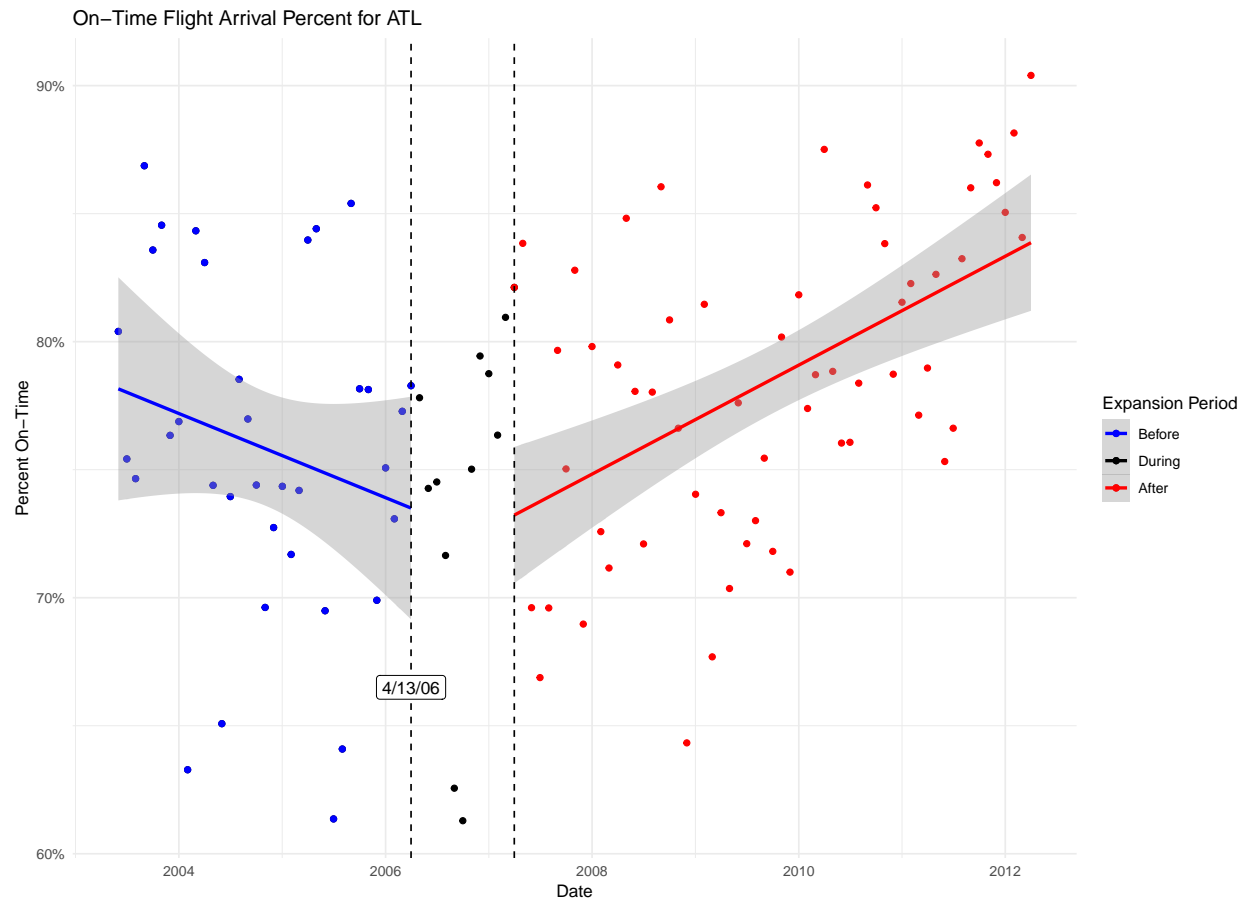


```
# Filtering to ATL:
standard_atl <- standard %>%
  filter(Facility == 'ATL')
  #mutate(exp_group = ifelse(Date >= "2006-04-01", "1", "0"))

# Percent On-Time Arrivals for ATL:
ggplot(standard_atl, aes(x=Date, y=Perc_On_Time_Gate_Arrs)) +
  geom_point(aes(color = "1")) +
  geom_point(data=subset(standard_atl, Date >= "2007-04-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_atl, Date >= "2007-04-01"), aes(color = "2"), method='lm', formula=y~x) +
  geom_point(data=subset(standard_atl, Date <= "2006-04-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_atl, Date <= "2006-04-01"), aes(color = "0"), method='lm', formula=y~x) +
  geom_vline(xintercept = as.numeric(as.Date("2007-04-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2006-04-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrival Percent for ATL") +
  scale_y_continuous(labels = function(x) paste0(x, "%")) +
  scale_color_manual(values = c("0" = "blue", "1" = "black", "2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Percent On-Time", color = "Expansion Period") +
  annotate("label", x = as.Date("2006-04-01"), y = 66.5, label = "4/13/06") +
  theme_minimal()
```



```
# Percent On-Time Arrivals for ATL zoomed version:
ggplot(standard_atl,aes(x=Date,y=Perc_On_Time_Gate_Arrs))+
  # Need this first subset to prevent plotting later years
  geom_point(data=subset(standard_atl, Date <= "2007-04-01"), aes(color = "1")) +
  # Now we only include up to a number we are comfortable with
  geom_point(data=subset(standard_atl, Date >= "2007-04-01" & Date <= "2012-04-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_atl, Date >= "2007-04-01" & Date <= "2012-04-01"), aes(color = "2"),
  geom_point(data=subset(standard_atl, Date <= "2006-04-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_atl, Date <= "2006-04-01"), aes(color = "0"), method='lm',formula=y
  geom_vline(xintercept = as.numeric(as.Date("2007-04-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2006-04-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrival Percent for ATL") +
  scale_y_continuous(labels = function(x) paste0(x, "%")) +
  scale_color_manual(values = c("0" = "blue","1" = "black","2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Percent On-Time", color = "Expansion Period") +
  annotate("label", x = as.Date("2006-04-01"), y = 66.5, label = "4/13/06") +
  theme_minimal()
```

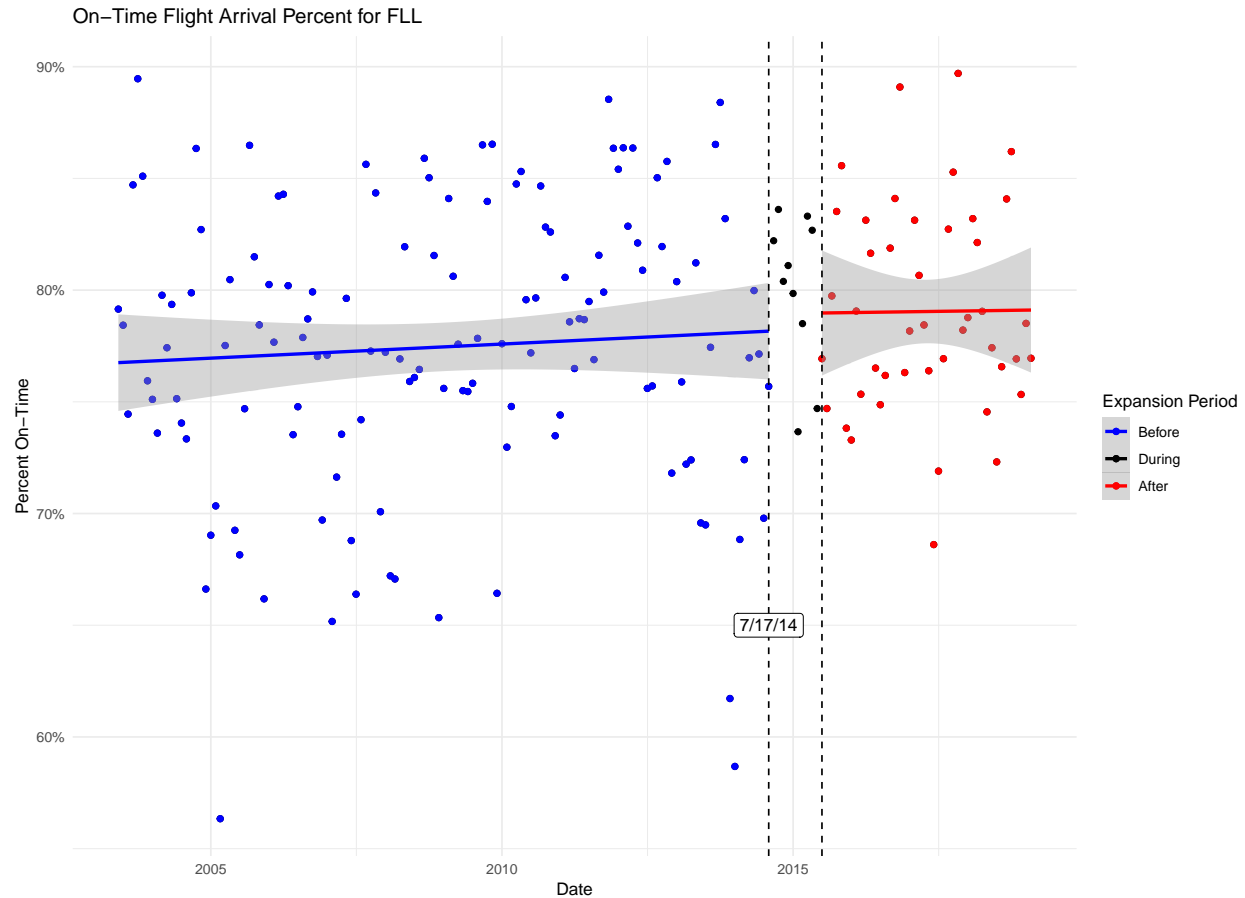


```
# Total On-Time Arrivals (OPS) fit to entire atl data
ggplot(standard_atl, aes(x=Date, y=On_Time_Arrivals)) +
  geom_point(aes(color = "1")) +
  geom_point(data=subset(standard_atl, Date >= "2007-04-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_atl, Date >= "2007-04-01"), aes(color = "2"), method='lm', formula=y~x) +
  geom_point(data=subset(standard_atl, Date <= "2006-04-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_atl, Date <= "2006-04-01"), aes(color = "0"), method='lm', formula=y~x) +
  geom_vline(xintercept = as.numeric(as.Date("2007-04-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2006-04-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrivals Total for ATL") +
  scale_color_manual(values = c("0" = "blue", "1" = "black", "2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Total On-Time Flight Arrivals", color = "Expansion Period") +
  annotate("label", x = as.Date("2006-04-01"), y = 20000, label = "4/13/06") +
  theme_minimal()
```

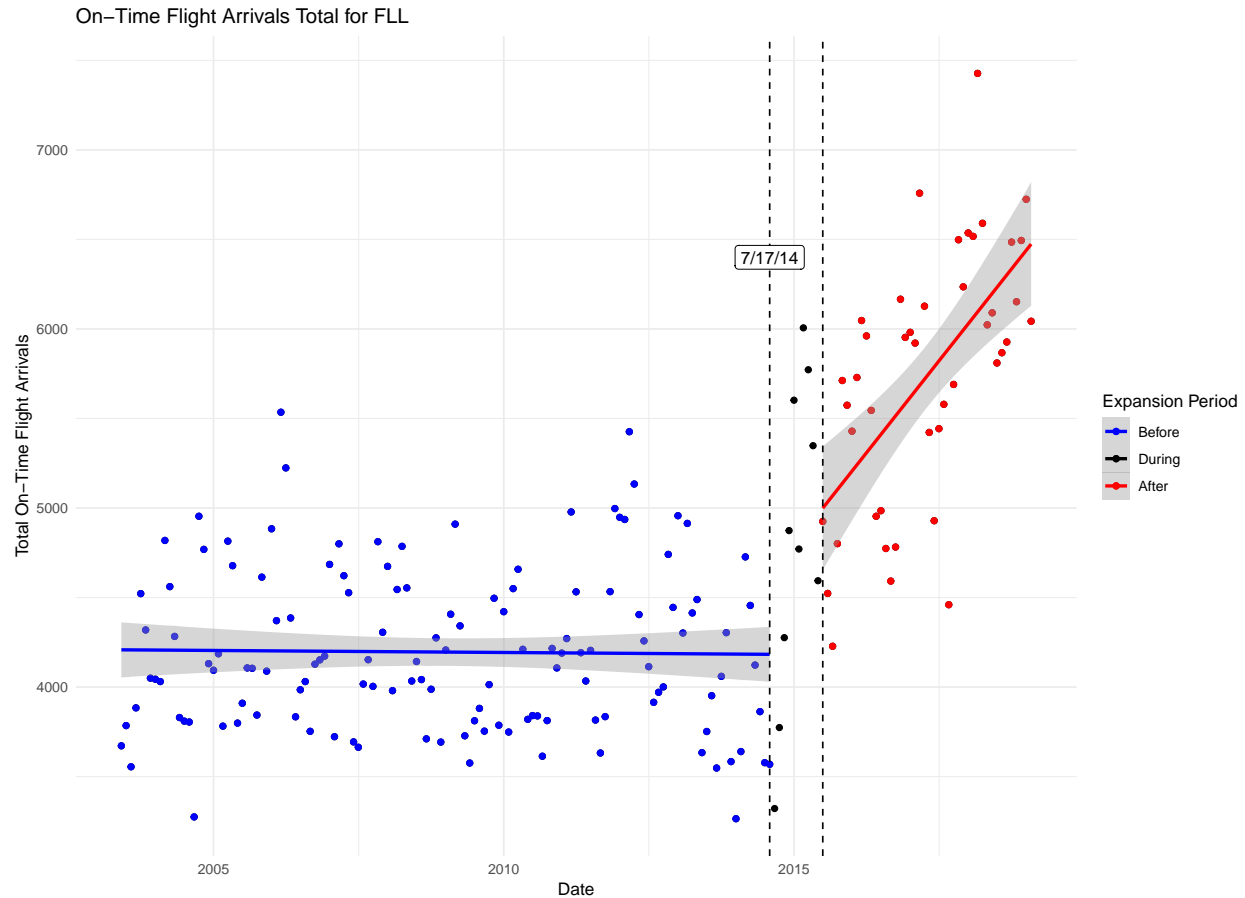


```
# Filtering to FLL:
standard_fll <- standard %>%
  filter(Facility == 'FLL')

# Percent On-Time Arrivals for FLL:
ggplot(standard_fll, aes(x=Date, y=Perc_On_Time_Gate_Arrs)) +
  geom_point(aes(color = "1")) +
  geom_point(data=subset(standard_fll, Date >= "2015-07-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_fll, Date >= "2015-07-01"), aes(color = "2"), method='lm', formula=y~x) +
  geom_point(data=subset(standard_fll, Date <= "2014-08-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_fll, Date <= "2014-08-01"), aes(color = "0"), method='lm', formula=y~x) +
  geom_vline(xintercept = as.numeric(as.Date("2015-07-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2014-08-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrival Percent for FLL") +
  scale_y_continuous(labels = function(x) paste0(x, "%")) +
  scale_color_manual(values = c("0" = "blue", "1" = "black", "2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Percent On-Time", color = "Expansion Period") +
  annotate("label", x = as.Date("2014-08-01"), y = 65, label = "7/17/14") +
  theme_minimal()
```

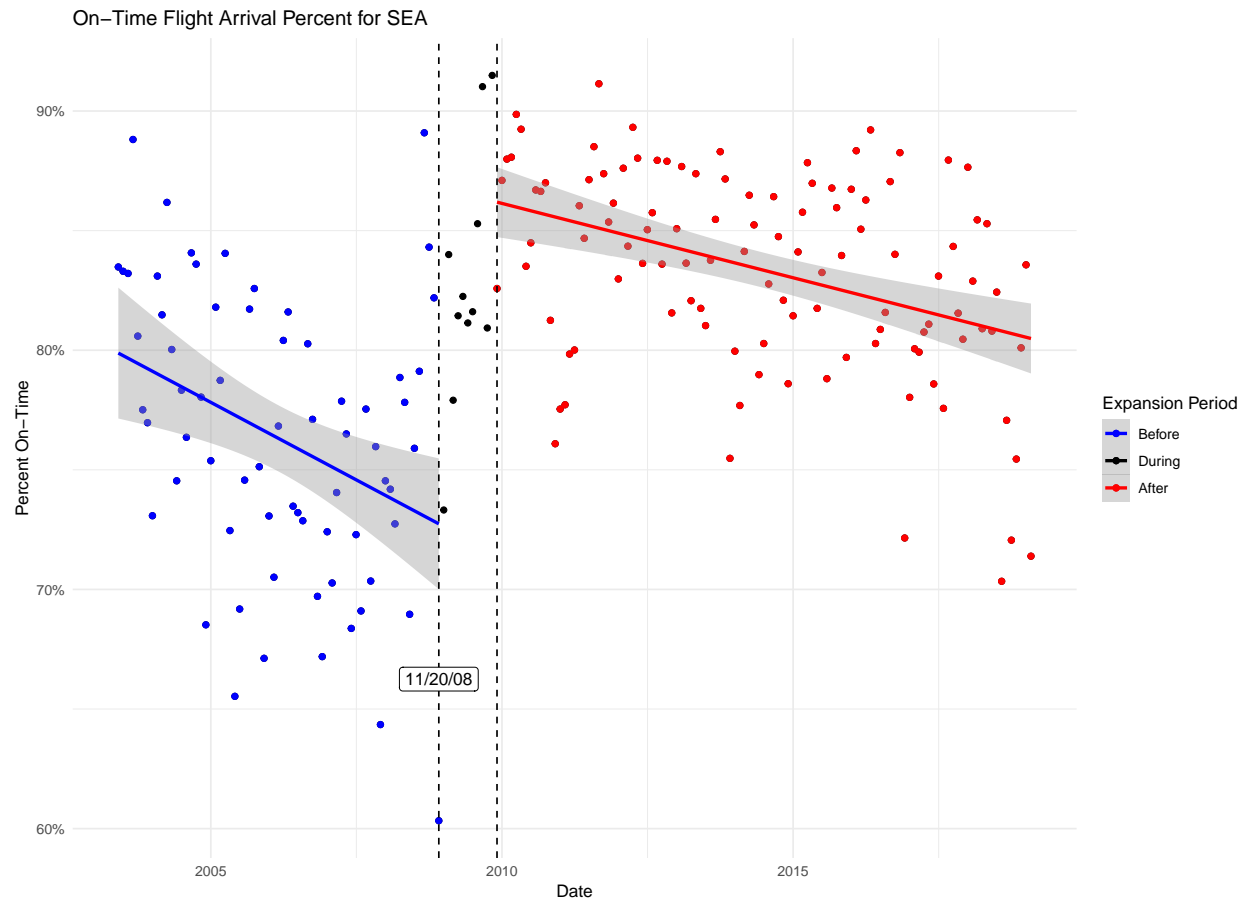


```
# Total On-Time Arrivals (OPS) fit to entire FLL data
ggplot(standard_fll, aes(x=Date, y=On_Time_Arrivals)) +
  geom_point(aes(color = "1")) +
  geom_point(data=subset(standard_fll, Date >= "2015-07-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_fll, Date >= "2015-07-01"), aes(color = "2"), method='lm', formula=y~x) +
  geom_point(data=subset(standard_fll, Date <= "2014-08-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_fll, Date <= "2014-08-01"), aes(color = "0"), method='lm', formula=y~x) +
  geom_vline(xintercept = as.numeric(as.Date("2015-07-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2014-08-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrivals Total for FLL") +
  scale_color_manual(values = c("0" = "blue", "1" = "black", "2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Total On-Time Flight Arrivals", color = "Expansion Period") +
  annotate("label", x = as.Date("2014-08-01"), y = 6400, label = "7/17/14") +
  theme_minimal()
```

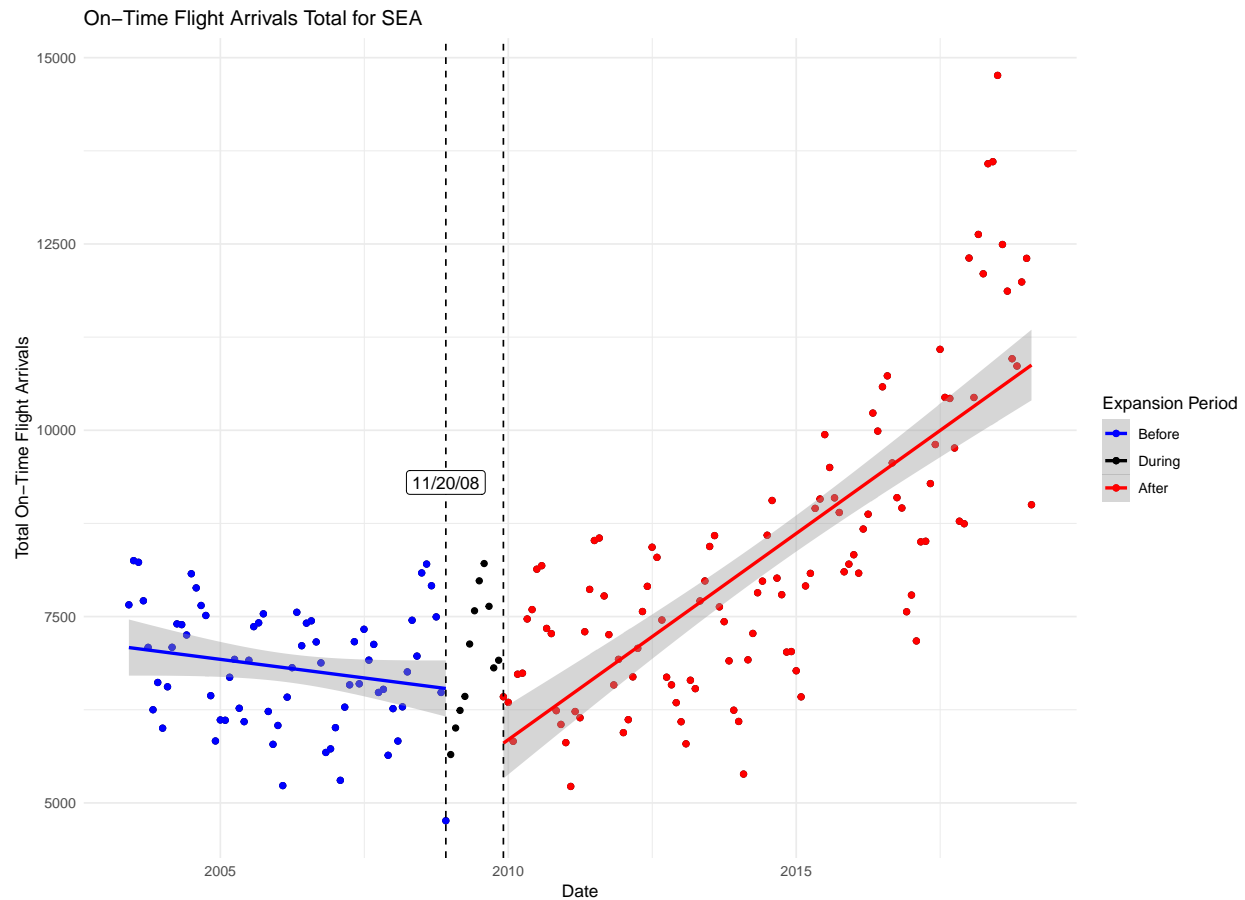


```
# Filtering to SEA:
standard_sea <- standard %>%
  filter(Facility == 'SEA')

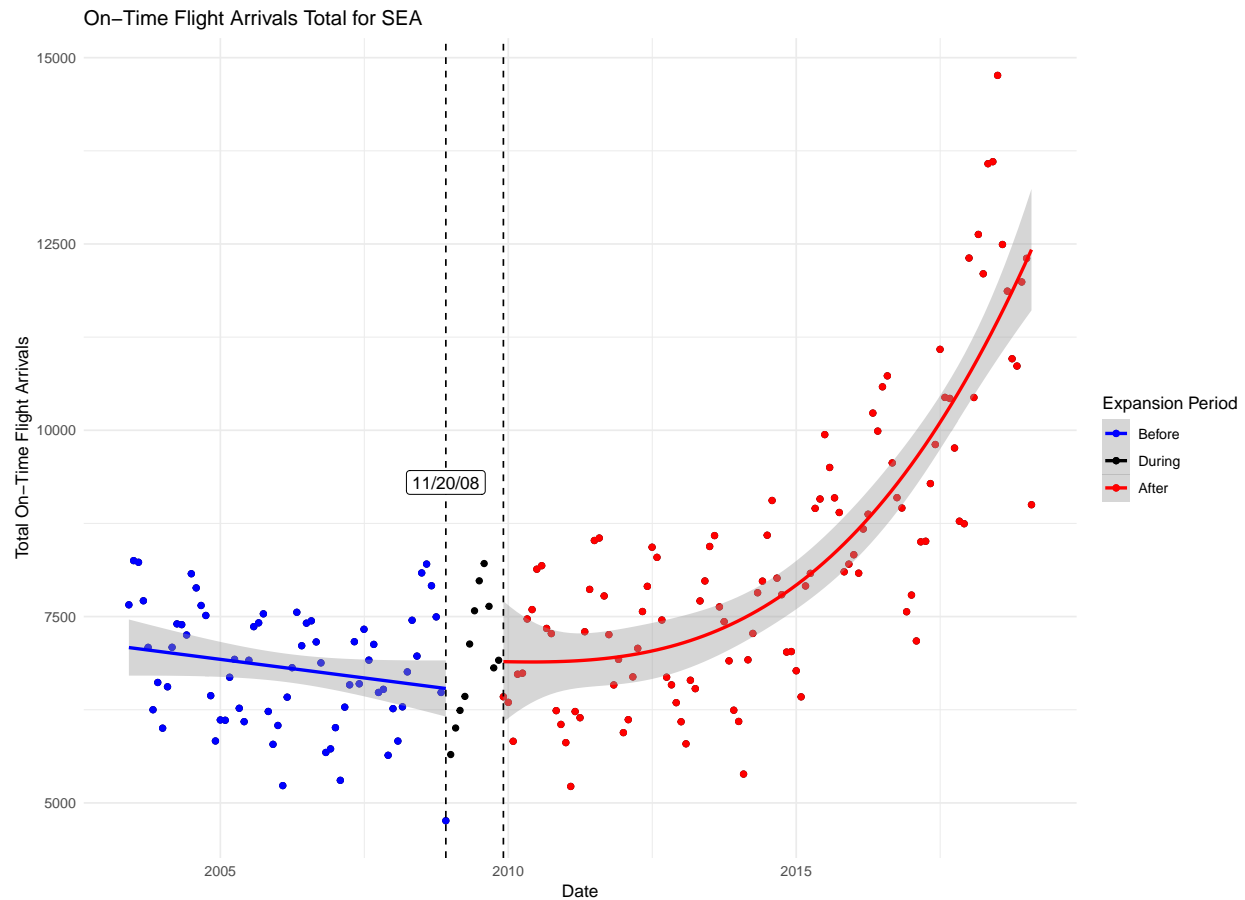
# Percent On-Time Arrivals for SEA:
ggplot(standard_sea, aes(x=Date, y=Perc_On_Time_Gate_Arrs)) +
  geom_point(aes(color = "1")) +
  geom_point(data=subset(standard_sea, Date >= "2009-12-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_sea, Date >= "2009-12-01"), aes(color = "2"), method='lm', formula=y~x) +
  geom_point(data=subset(standard_sea, Date <= "2008-12-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_sea, Date <= "2008-12-01"), aes(color = "0"), method='lm', formula=y~x) +
  geom_vline(xintercept = as.numeric(as.Date("2009-12-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2008-12-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrival Percent for SEA") +
  scale_y_continuous(labels = function(x) paste0(x, "%")) +
  scale_color_manual(values = c("0" = "blue", "1" = "black", "2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Percent On-Time", color = "Expansion Period") +
  annotate("label", x = as.Date("2008-12-01"), y = 66.25, label = "11/20/08") +
  theme_minimal()
```



```
# Total On-Time Arrivals (OPS) fit to entire SEA data
ggplot(standard_sea, aes(x=Date, y=On_Time_Arrivals)) +
  geom_point(aes(color = "1")) +
  geom_point(data=subset(standard_sea, Date >= "2009-12-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_sea, Date >= "2009-12-01"), aes(color = "2"), method='lm', formula=y~x) +
  geom_point(data=subset(standard_sea, Date <= "2008-12-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_sea, Date <= "2008-12-01"), aes(color = "0"), method='lm', formula=y~x) +
  geom_vline(xintercept = as.numeric(as.Date("2009-12-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2008-12-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrivals Total for SEA") +
  scale_color_manual(values = c("0" = "blue", "1" = "black", "2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Total On-Time Flight Arrivals", color = "Expansion Period") +
  annotate("label", x = as.Date("2008-12-01"), y = 9300, label = "11/20/08") +
  theme_minimal()
```

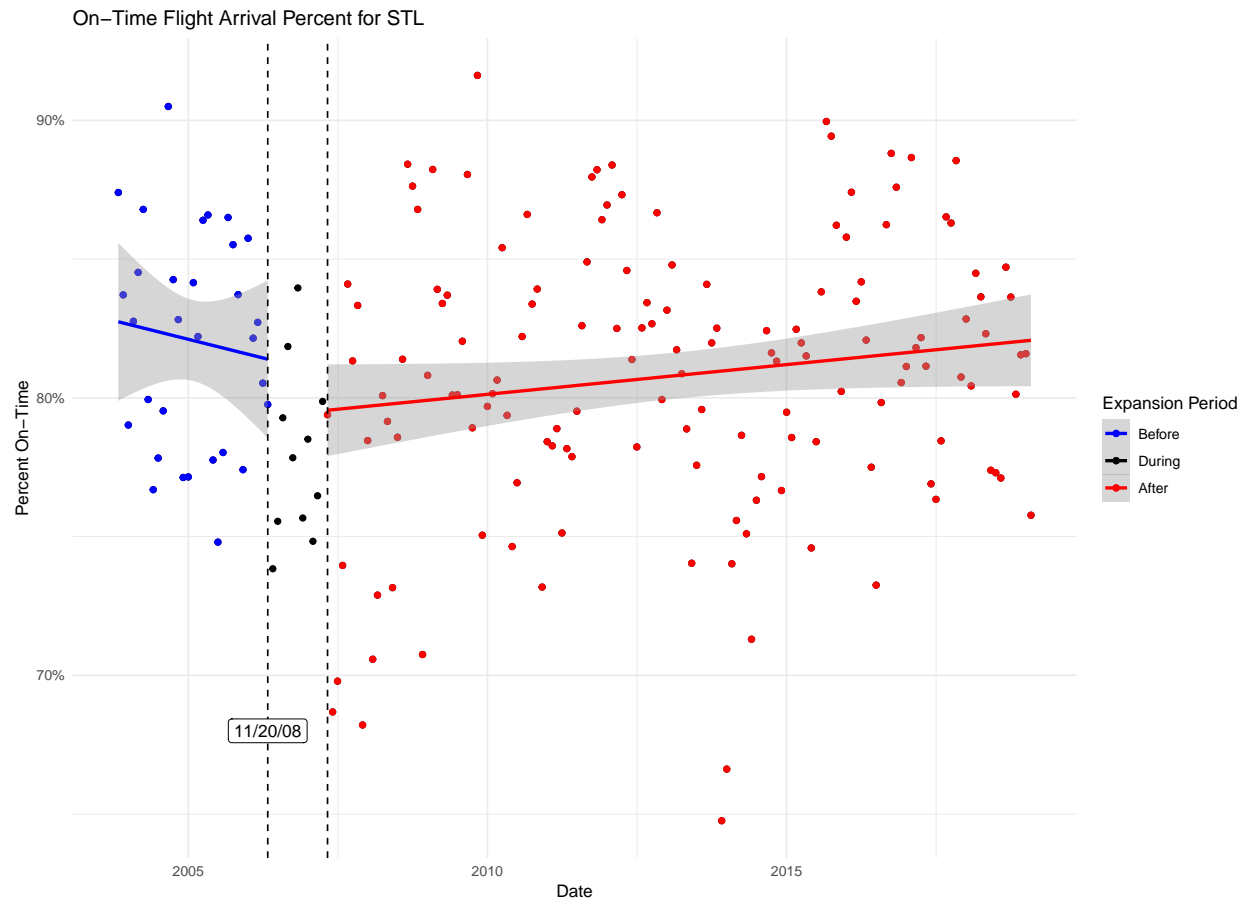



```
# Total On-Time Arrivals (OPS) fit to entire SEA data w/ cubic:
ggplot(standard_sea, aes(x=Date, y=On_Time_Arrivals)) +
  geom_point(aes(color = "1")) +
  geom_point(data=subset(standard_sea, Date >= "2009-12-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_sea, Date >= "2009-12-01"), aes(color = "2"), method='lm', formula=y ~ x) +
  geom_point(data=subset(standard_sea, Date <= "2008-12-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_sea, Date <= "2008-12-01"), aes(color = "0"), method='lm', formula=y ~ x) +
  geom_vline(xintercept = as.numeric(as.Date("2009-12-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2008-12-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrivals Total for SEA") +
  scale_color_manual(values = c("0" = "blue", "1" = "black", "2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Total On-Time Flight Arrivals", color = "Expansion Period") +
  annotate("label", x = as.Date("2008-12-01"), y = 9300, label = "11/20/08") +
  theme_minimal()
```

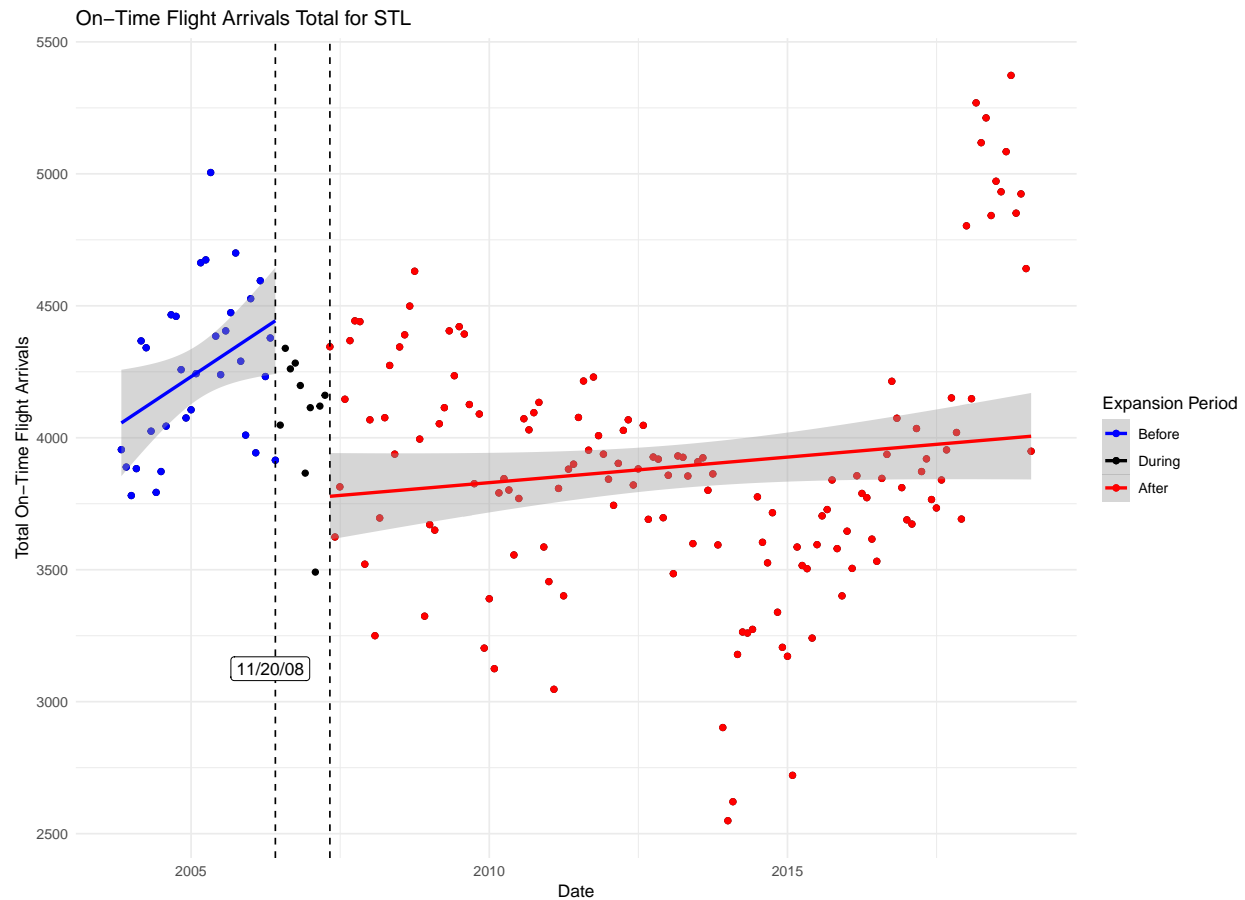


```
# Filtering to STL:
standard_stl <- standard %>%
  filter(Facility == 'STL') %>%
  filter(Date >= "2003-11-01")

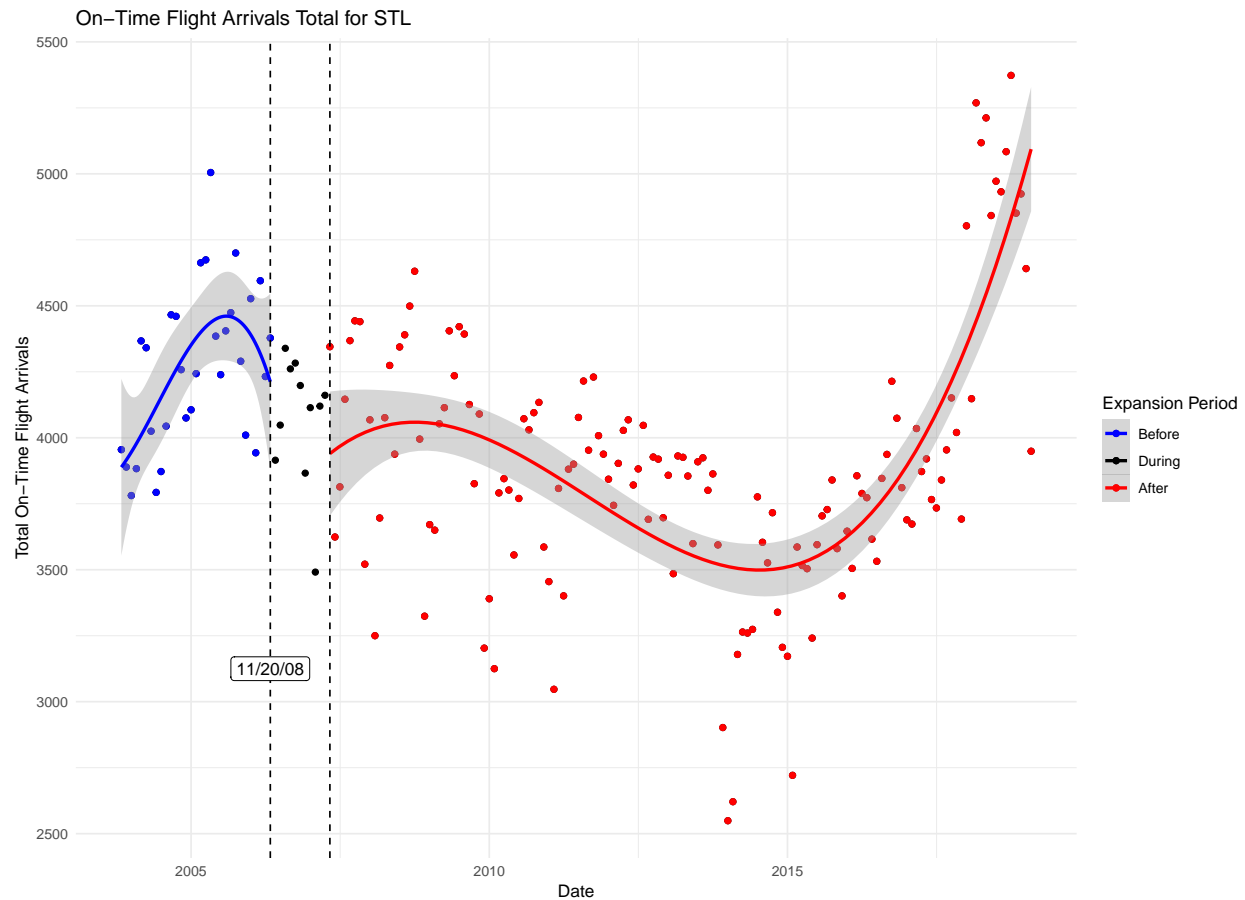
# Percent On-Time Arrivals for STL:
ggplot(standard_stl, aes(x=Date, y=Perc_On_Time_Gate_Arrs)) +
  geom_point(aes(color = "1")) +
  geom_point(data=subset(standard_stl, Date >= "2007-05-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_stl, Date >= "2007-05-01"), aes(color = "2"), method='lm', formula=y~x) +
  geom_point(data=subset(standard_stl, Date <= "2006-05-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_stl, Date <= "2006-05-01"), aes(color = "0"), method='lm', formula=y~x) +
  geom_vline(xintercept = as.numeric(as.Date("2007-05-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2006-05-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrival Percent for STL") +
  scale_y_continuous(labels = function(x) paste0(x, "%")) +
  scale_color_manual(values = c("0" = "blue", "1" = "black", "2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Percent On-Time", color = "Expansion Period") +
  annotate("label", x = as.Date("2006-05-01"), y = 68, label = "11/20/08") +
  theme_minimal()
```



```
# Total On-Time Arrivals (OPS) fit to entire STL data
ggplot(standard_stl, aes(x=Date, y=On_Time_Arrivals)) +
  geom_point(aes(color = "1")) +
  geom_point(data=subset(standard_stl, Date >= "2007-05-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_stl, Date >= "2007-05-01"), aes(color = "2"), method='lm', formula=y~x) +
  geom_point(data=subset(standard_stl, Date <= "2006-06-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_stl, Date <= "2006-06-01"), aes(color = "0"), method='lm', formula=y~x) +
  geom_vline(xintercept = as.numeric(as.Date("2007-05-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2006-06-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrivals Total for STL") +
  scale_color_manual(values = c("0" = "blue", "1" = "black", "2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Total On-Time Flight Arrivals", color = "Expansion Period") +
  annotate("label", x = as.Date("2006-05-01"), y = 3125, label = "11/20/08") +
  theme_minimal()
```



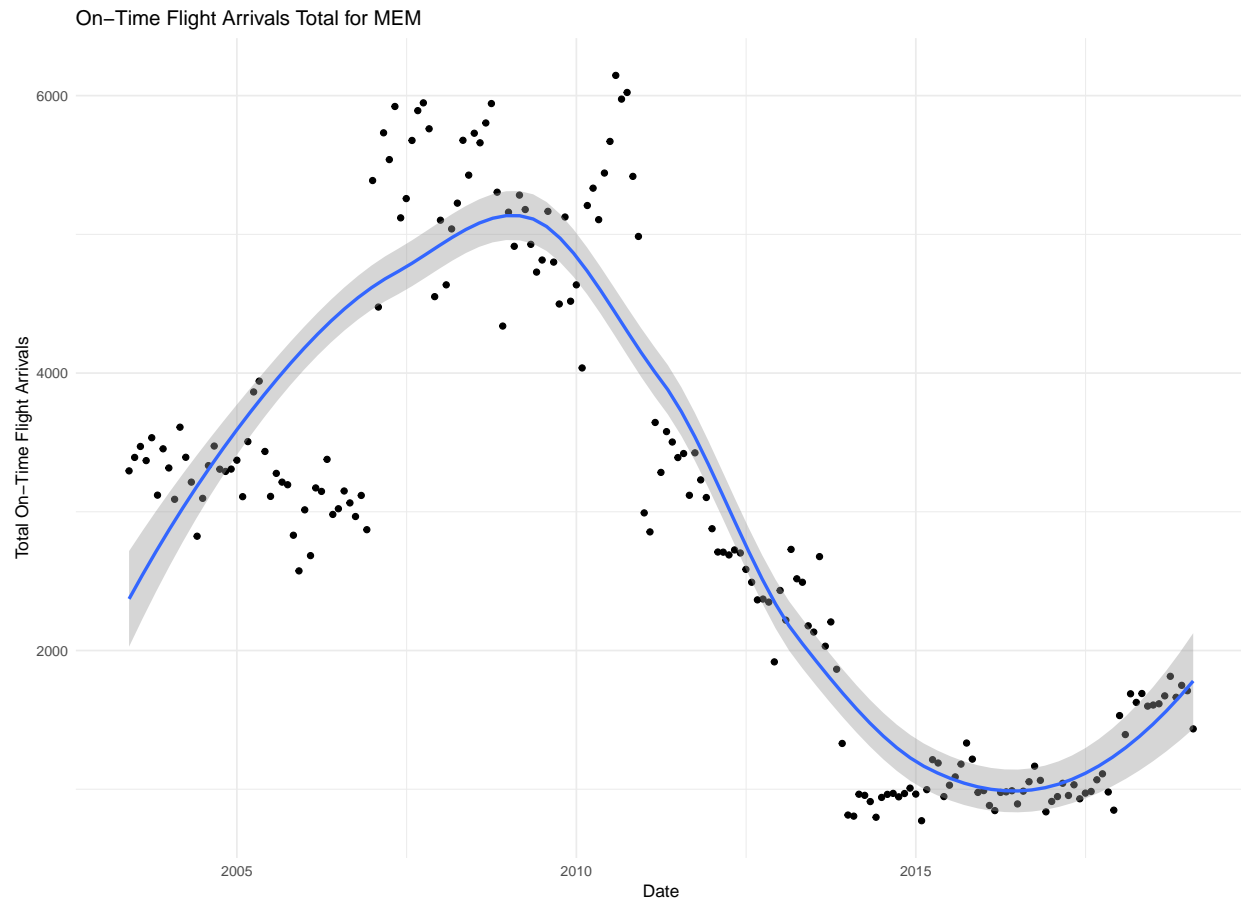
```
# Total On-Time Arrivals (OPS) fit to entire STL data w/ cubic
ggplot(standard_stl,aes(x=Date,y=On_Time_Arrivals))+
  geom_point(aes(color = "1")) +
  geom_point(data=subset(standard_stl, Date >= "2007-05-01"), aes(color = "2")) +
  geom_smooth(data=subset(standard_stl, Date >= "2007-05-01"), aes(color = "2"), method='lm',formula=y ~ x) +
  geom_point(data=subset(standard_stl, Date <= "2006-05-01"), aes(color = "0")) +
  geom_smooth(data=subset(standard_stl, Date <= "2006-05-01"), aes(color = "0"), method='lm',formula=y ~ x) +
  geom_vline(xintercept = as.numeric(as.Date("2007-05-01")), linetype='dashed') +
  geom_vline(xintercept = as.numeric(as.Date("2006-05-01")), linetype='dashed') +
  ggtitle(label = "On-Time Flight Arrivals Total for STL") +
  scale_color_manual(values = c("0" = "blue","1" = "black","2" = "red"),
    labels = c("Before", "During", "After")) +
  labs(y = "Total On-Time Flight Arrivals", color = "Expansion Period") +
  annotate("label", x = as.Date("2006-05-01"), y = 3125, label = "11/20/08") +
  theme_minimal()
```



```
# Filtering to MEM:
standard_mem <- standard %>%
  filter(Facility == 'MEM')

# Total On-Time Arrivals (OPS) for MEM:
ggplot(standard_mem, aes(x=Date, y=On_Time_Arrivals)) +
  geom_point() +
  geom_smooth() +
  ggtitle(label = "On-Time Flight Arrivals Total for MEM") +
  labs(y = "Total On-Time Flight Arrivals") +
  theme_minimal()

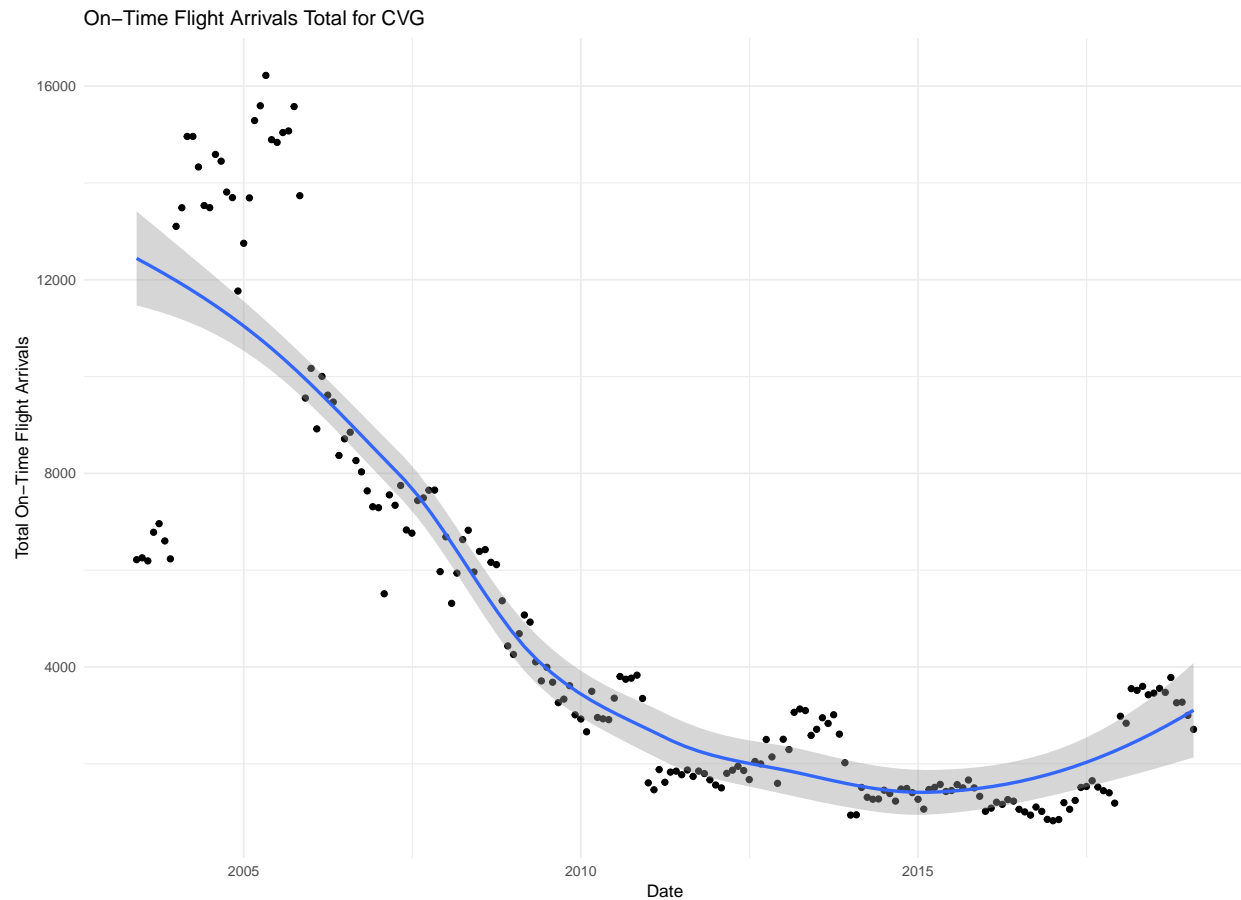
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
# Filtering to CVG:
standard_cvg <- standard %>%
  filter(Facility == 'CVG')

# Total On-Time Arrivals (OPS) for CVG:
ggplot(standard_cvg, aes(x=Date, y=On_Time_Arrivals)) +
  geom_point() +
  geom_smooth() +
  ggtitle(label = "On-Time Flight Arrivals Total for CVG") +
  labs(y = "Total On-Time Flight Arrivals") +
  theme_minimal()

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

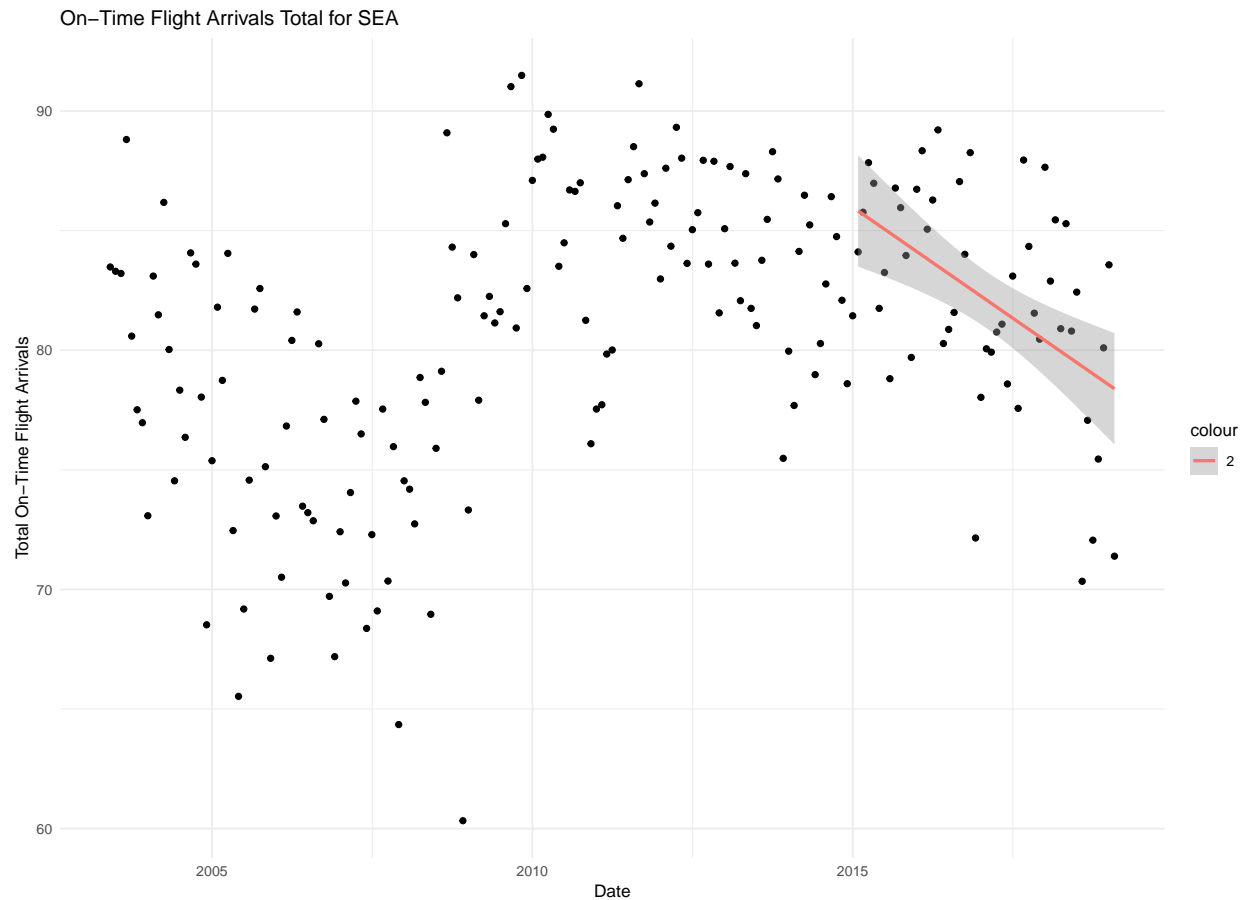


Coefficient Comparisons

Example of WTF we're doing:

```
# Fitting linear model to all airports for last 4 years:
four_year_standard <- standard %>%
  filter(Date >= "2015-02-01")

# Percent On-Time Arrivals (OPS) for SEA:
ggplot(standard_sea, aes(x=Date, y=Perc_On_Time_Gate_Arrs)) +
  geom_point() +
  geom_smooth(data=subset(standard_sea, Date >= "2015-02-01"), aes(color = "2"), method='lm', formula=y~x) +
  ggtitle(label = "On-Time Flight Arrivals Total for SEA") +
  labs(y = "Total On-Time Flight Arrivals") +
  theme_minimal()
```

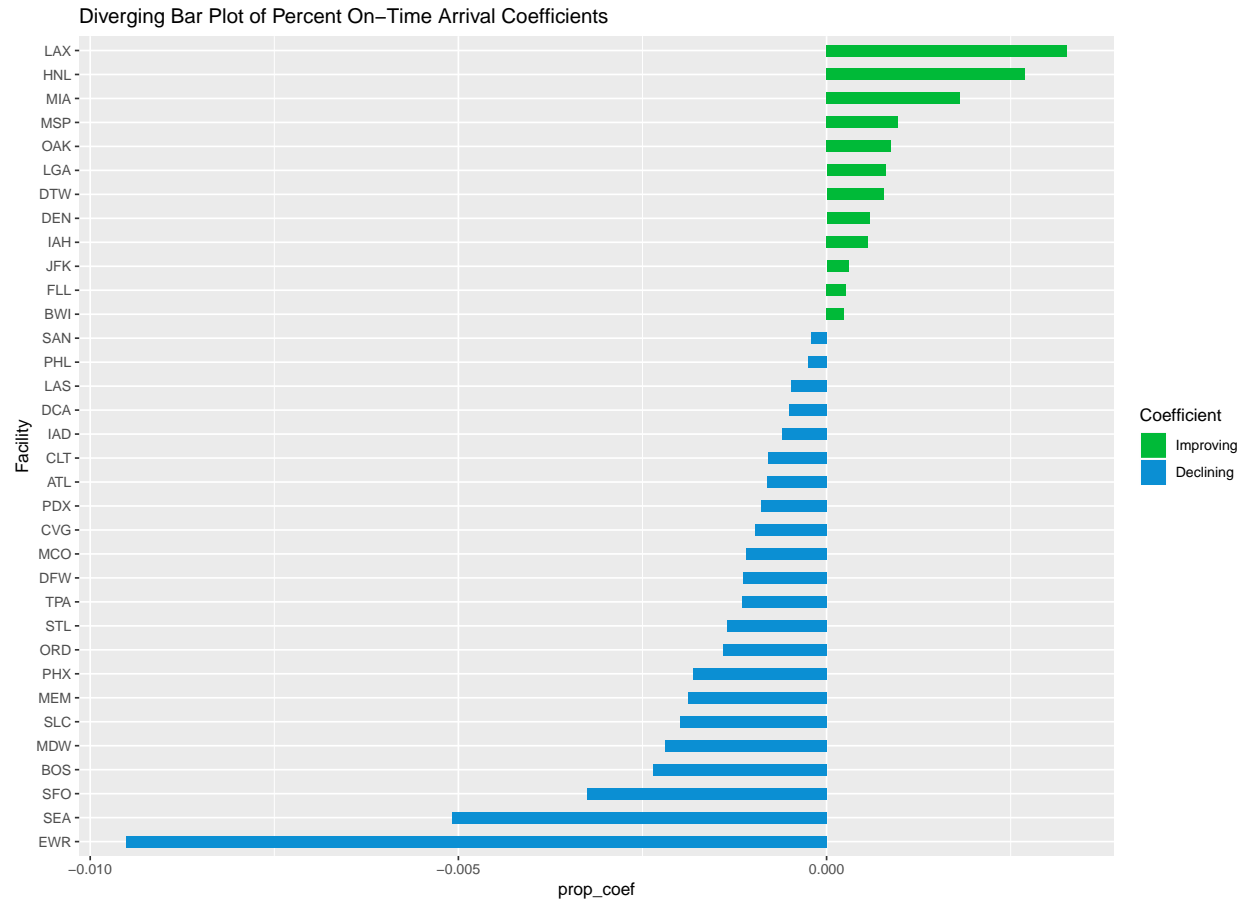


```
prop_on_time_fit <- lmList(Perc_On_Time_Gate_Arrs ~ Date | Facility, data=four_year_standard, pool=FALSE)

prop_on_time_coefs <- coef(prop_on_time_fit) %>%
  tibble::rownames_to_column() %>%
  rename(Facility = rowname, prop_coef = Date) %>%
  select(Facility, prop_coef) %>%
  mutate(Date = prop_coef*30.42) %>%
  arrange(desc(prop_coef)) %>%
  mutate(prop_coef_type = ifelse(prop_coef <= 0, "below", "above"))

prop_on_time_coefs <- prop_on_time_coefs[order(prop_on_time_coefs$prop_coef), ] #Ascending sort on Z Score
prop_on_time_coefs$Facility <- factor(prop_on_time_coefs$Facility, levels = prop_on_time_coefs$Facility)

ggplot(prop_on_time_coefs, aes(x=Facility, y=prop_coef, label=prop_coef)) +
  geom_bar(stat='identity', aes(fill=prop_coef_type), width=.5) +
  scale_fill_manual(name="Coefficient",
    labels = c("Improving", "Declining"),
    values = c("above"="#00ba38", "below"="#0b8fd3")) +
  labs(title= "Diverging Bar Plot of Percent On-Time Arrival Coefficients") +
  coord_flip()
```

```
four_year_standardized_standard <- standard %>%
  filter(Date >= "2015-02-01") %>%
  group_by(Facility) %>%
  mutate(std.On_Time_Arrivals = scale(On_Time_Arrivals))

arr_on_time_fit <- lmList(std.On_Time_Arrivals ~ Date | Facility, data=four_year_standardized_standard,

arr_on_time_coefs <- coef(arr_on_time_fit) %>%
  tibble::rownames_to_column() %>%
  rename(Facility = rowname, arr_coef = Date) %>%
  select(Facility, arr_coef) %>%
  mutate(arr_coef = arr_coef*30.42) %>%
  arrange(desc(arr_coef)) %>%
  mutate(arr_coef_type = ifelse(arr_coef <= 0, "below", "above"))

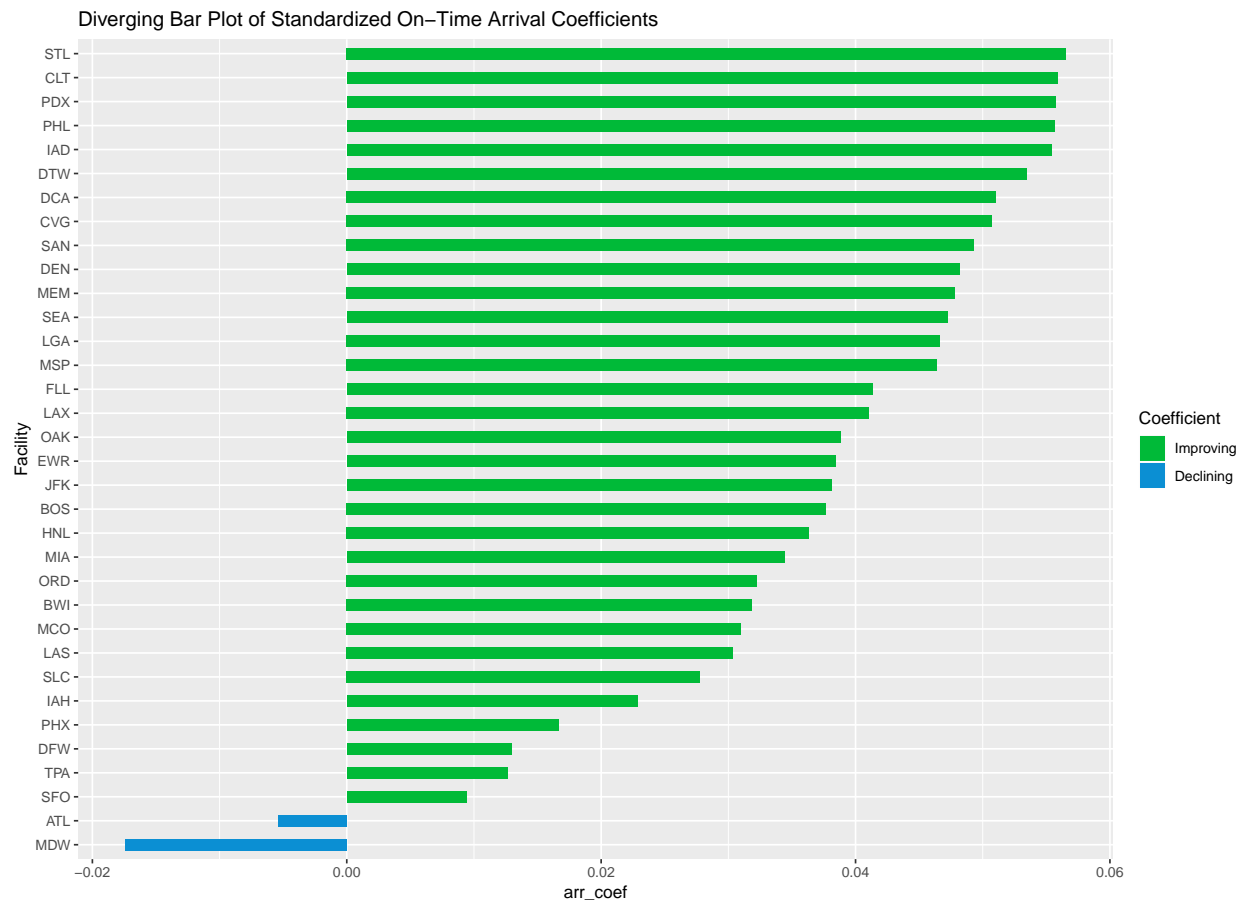
arr_on_time_coefs <- arr_on_time_coefs[order(arr_on_time_coefs$arr_coef), ] #Ascending sort on Z Score
arr_on_time_coefs$Facility <- factor(arr_on_time_coefs$Facility, levels = arr_on_time_coefs$Facility)

ggplot(arr_on_time_coefs, aes(x=Facility, y=arr_coef, label=arr_coef)) +
  geom_bar(stat='identity', aes(fill=arr_coef_type), width=.5) +
  scale_fill_manual(name="Coefficient",
    labels = c("Improving", "Declining"),
```

```

    values = c("above"="#00ba38", "below"="#0b8fd3")) +
labs(title= "Diverging Bar Plot of Standardized On-Time Arrival Coefficients") +
coord_flip()

```

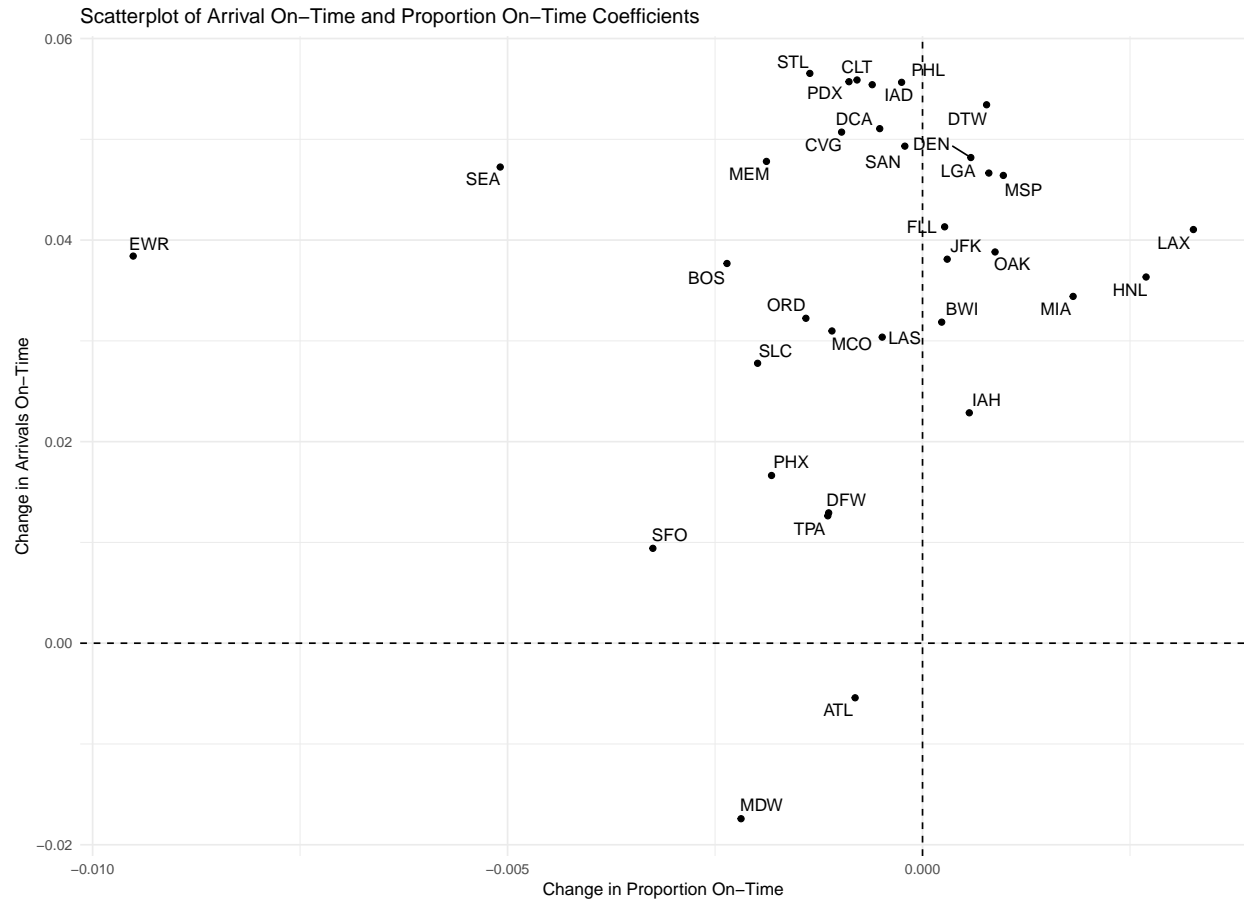


```

merged_coefs <- merge(prop_on_time_coefs, arr_on_time_coefs, by=c("Facility"))

ggplot(merged_coefs, aes(x = prop_coef, y = arr_coef)) +
  geom_vline(xintercept = 0, linetype='dashed') +
  geom_hline(yintercept = 0, linetype='dashed') +
  geom_point() +
  ggrepel::geom_text_repel(aes(label = Facility)) +
  ggtitle(label = "Scatterplot of Arrival On-Time and Proportion On-Time Coefficients") +
  labs(x = "Change in Proportion On-Time", y = "Change in Arrivals On-Time") +
  theme_minimal()

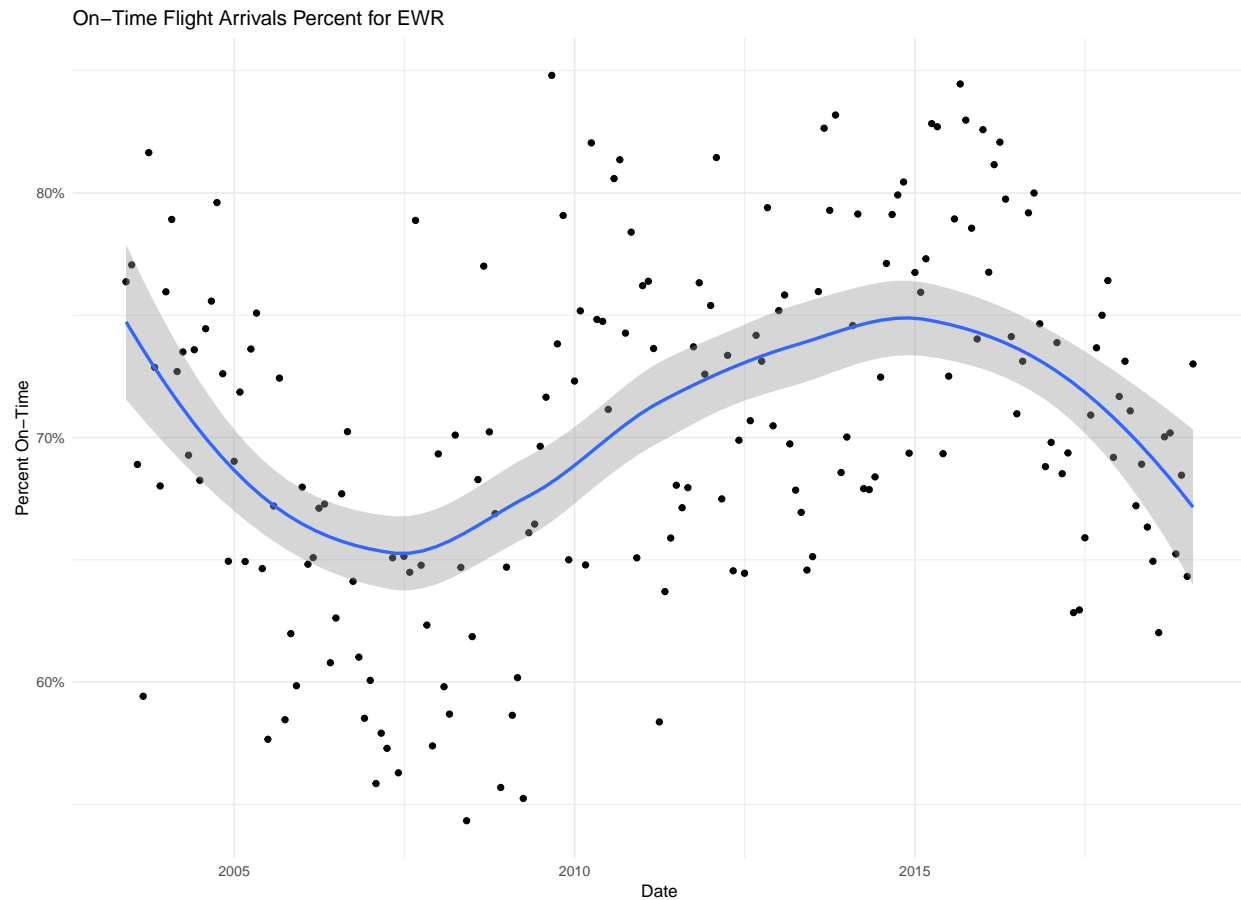
```



```
# Let's look at WTF is happening at EWR:
standard_ewr <- standard %>%
  filter(Facility == 'EWR')

# Percent on time arrivals for EWR:
ggplot(standard_ewr, aes(x=Date, y=Perc_On_Time_Gate_Arrs)) +
  geom_point() +
  geom_smooth() +
  ggtitle(label = "On-Time Flight Arrivals Percent for EWR") +
  scale_y_continuous(labels = function(x) paste0(x, "%")) +
  labs(y = "Percent On-Time") +
  theme_minimal()

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
ggplot(standard_ewr, aes(x=Date, y=On_Time_Arrivals)) +  
  geom_point() +  
  geom_smooth() +  
  ggtitle(label = "On-Time Flight Arrival Total for EWR") +  
  labs(y = "Total On-Time Arrivals") +  
  theme_minimal()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

