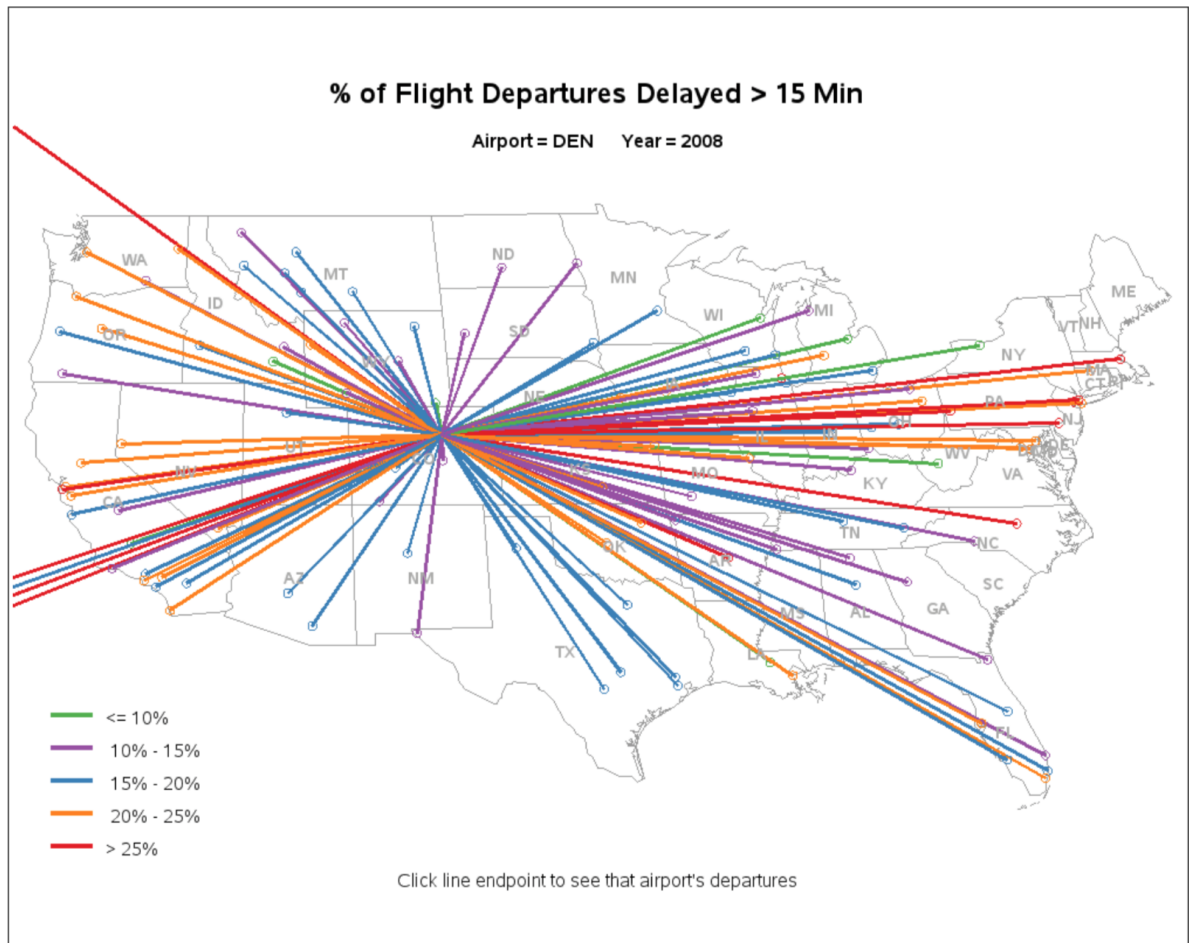


מעבדה ראשונה - ניתוח נתוני טיסות

עדי זיו וספיר חן-ציון

שאלה ראשונה - ניתוח גרפים קיימים

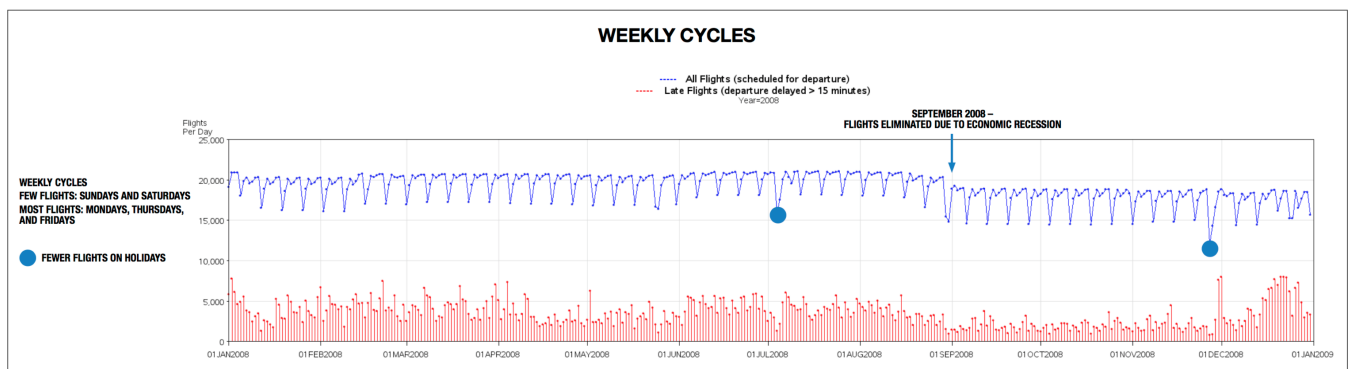


הגרף מציג את סיפורו של נמל התעופה דנוור, בקולורדו, בשנת 2008. את הסיפור שמספר הגרף ניתן לחלק לשני רבדים: ברובד הפשוט ביותר מוצגת תמונה של כל היעדים שיוצאים מדנוור למדינות ארצות הברית. כאשר, בעבור כל יעד, מוצג לצופה אחוז הטיסות המתעכבות בהמראה עיכוב משמעותי (מעל 15 דקות). ברובד המורכב יותר והנתון לפרשנות, ניתן לראות תמונה המספרת כי "הסתברות" לעיכוב משמעותי גדולה יותר בעבור יעדים רחוקים יותר (כך לדוגמה ההסתברות לעיכוב בטיסות לאלסקה, הוואי, הונולולו, או החוף המזרחי, גדול משמעותית מההסתברות לעיכוב במונטנה, ויומינג, ניו מקסיקו, טקסס ועוד).

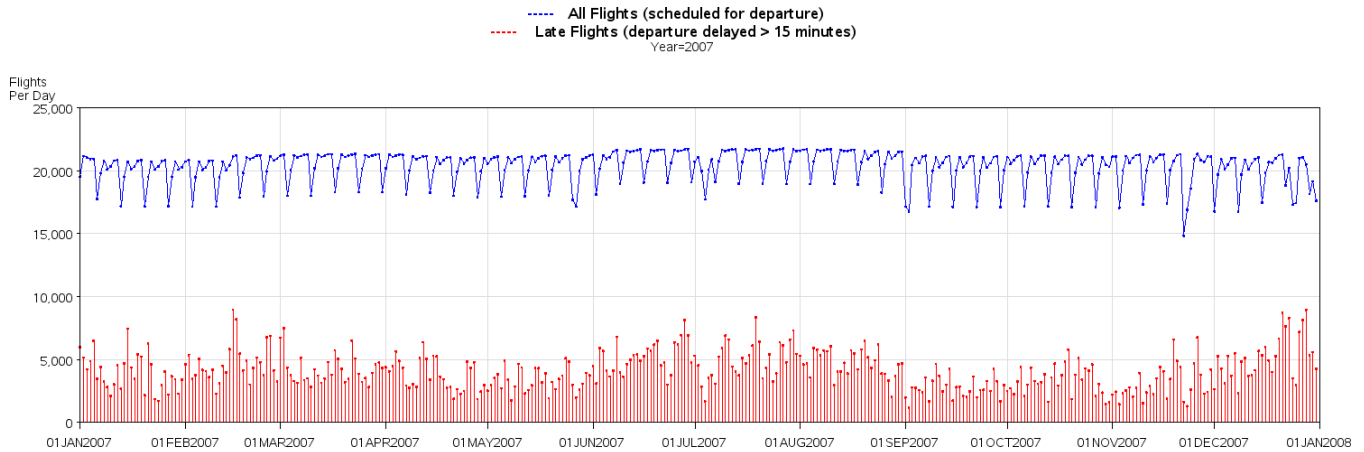
לדעתנו, ישנו כישלון העולה מן הגרף בלספר סיפור מורכב יותר. בעקבות הכישלון והתוצאה היצריתית עולות מספר שאלות נוספות: הראשונה, שעולה כמעט באופן מיידי: מהן כמות הטיסות לכל יעד? שאלה זו פותחת דלת לסוגיה האם קיים קשר בין כמות הטיסות לבין העיכובים? שאלה נוספת שעולה ותעזור למקד את הסיפור שאותו הגרף מספר מתייחסת אל שדה התעופה בדנוור: איך מתנהל שדה התעופה בדנוור אל מול שדות תעופה אחרים? האם יש בו טיסות רבות יותר ולכן עיכובים רבים יותר? האם ניתן לראות דפוסים דומים של הסתברות גבוהה יותר לעיכובים גם בשדות תעופה אחרים? פתרון שאלה זו שבא לידי ביטוי בהוספת גרפים דומים של נמלי תעופה אחרים, מעיד על יכולת של הגרף הנכחי לספר סיפור מעמיק ומבוסס יותר.

נציע לספר את הסיפור בצורה קצת שונה: ברמה הגרפית, קווים דקים יותר, וייצוג אחוז הטיסות המתעכבות באמצעות צבע הנקודה במקום, ייצוג מצומצם יהפוך את הצפייה בגרף לנעימה וברורה יותר. ברמה הסיפורית, נציע להציג באמצעות גודל הנקודות את כמות הטיסות היוצאות ליעד, זאת בכדי לספק עוד מידע לצופה. כמו כן, כדי לענות על השאלה – האם ההסתברות לעיכוב גדלה בעבור יעדים רחוקים יותר? נציע להוסיף על הגרף טבעות ולהציג את ההסתברות לעיכוב משמעותי בעבור כל טבעת. ניתן גם להציג מספר גרפים במקביל של שדות תעופה שונים.

TEMPORAL EFFECTS



במבט ראשוני על הגרף ניתן לראות שהסיפור העיקרי שמסופר הוא כמות הטיסות אל מול כמות האיחורים ביום, במשך שנת 2008. את הסיפור של הגרף ניתן לבחון בשני רבדים, שבועי ושנתי. ברמה השבועית ניתן לראות כי בכל שבוע בימי ראשון ושבת מתוכננות פחות טיסות מבשר השבוע. כמו כן ניתן לראות, בצורה חלקית בלבד, כי בימים אלו כמות האיחורים עולה. נציין כי מסקנה זו אינה חד משמעית לאורך הגרף. ברמה השנתית מנסה הגרף לספר שבמהלך החגים כמות הטיסות יורדת זמנית ועולה מיד לאחר החג בחזרה. כמו גם, שבעקבות המשבר הכלכלי בשנת 2008 ירדו כמות הטיסות בצורה משמעותית – נתייחס לנקודה זאת בהמשך כאשר נציע שיפורים. לדעתנו, צורת המענה לשאלות הללו היא חלקית בלבד ואינה מספקת. כשליו העיקריים של הגרף נובעים מכך שברמה היומית והשנתית קשה לרדת לפרטים הקטנים הן בשל כמות הנתונים, צורת הצגתם ופריסתם. כשל זה בולט במיוחד בכך שכמות הטיסות המתוכננות מתוארת בצורה רציפה בעוד כמות הטיסות המתעכבות מתוארת בצורה בדידה. ברובד השנתי מתגלה כשל עיקרי שבא לידי ביטוי בחוסר התייחסות לאירועים משמעותיים הפועלים במחזוריות במהלך השנה. כך למשל מצוינים על הגרף רק חלק מהחגים. לדוגמה, סביב חג המולד או ליל כל הקדושים לא ניתן לראות שינוי דומה. הצגה זו מעוותת את התמונה וגורמת להסקת מסקנות שאינן נכונות בהכרח. כאמור, מצוין בגרף תחילת המשבר הכלכלי בספטמבר 2008 בה ירדו כמות הטיסות בצורה משמעותית, הסבר נוסף שניתן היה להעלות על הדעת הוא פתיחת שנת הלימודים. כפי שמובא, מבט על הגרף בשנים אחרות מראה לנו שהתופעה חוזרת על עצמה ושההסבר המובא כייחודי לשנת 2008 שגוי.



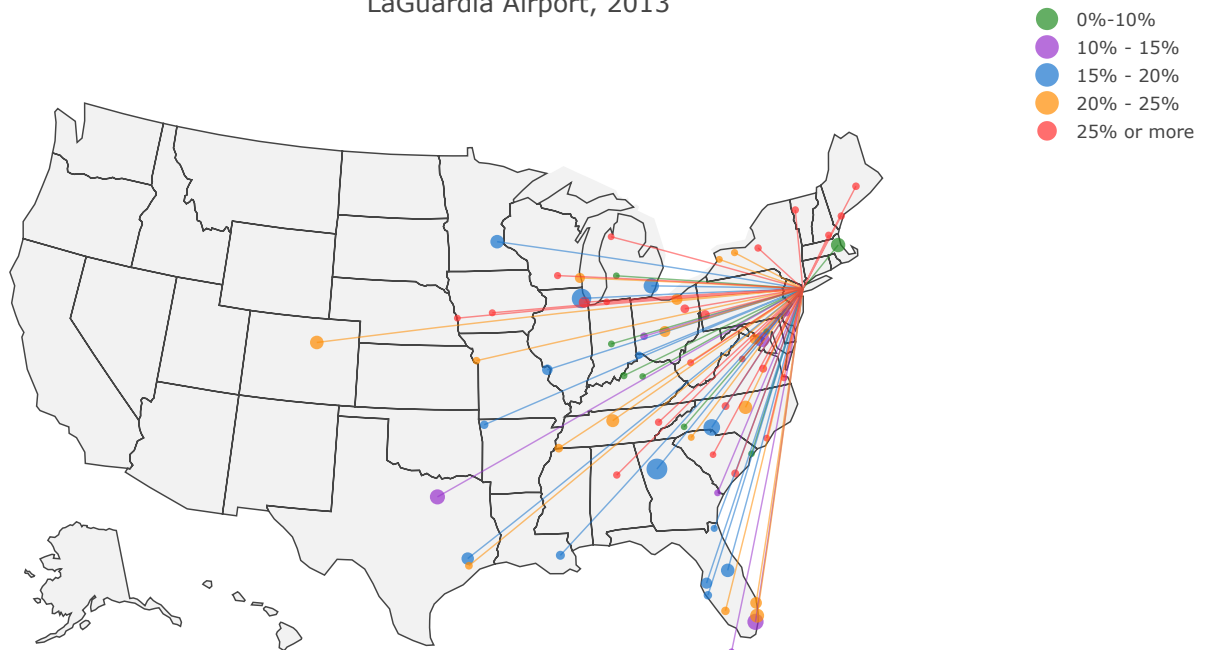
כשל נוסף הינו חוסר אינפורמציה: האם המידע המתואר מייצג את כלל ארצות הברית או נמל תעופה מסוים? האם הוא כולל טיסות פנים וחוצי כאחד?

על מנת לשפר את הגרף נציע מספר נקודות עיקריות: הראשונה בהם, תצוגה רציפה יותר, תוך הוספת ממד נוסף של צבע כדי לייצג באחוזים את כמות הטיסות המתעכבות. בתצורתו הנוכחית של הגרף לא ניתן להשוות בצורה טובה בין העיכובים בחודשים בהן ישנן טיסות מועטות יותר או רבות.

כמו כן, במידה והטענה היא שסביב חגים יש ירידה בכמות הטיסות הכוללת, יש להציג את כלל החגים. על מנת להשלים את התמונה בצורה הרחבה ביותר מומלץ להראות גם שנים נוספות ולסמן אירועים מחזוריים רלוונטיים נוספים.

שאלה שניה - שחזור גרפים

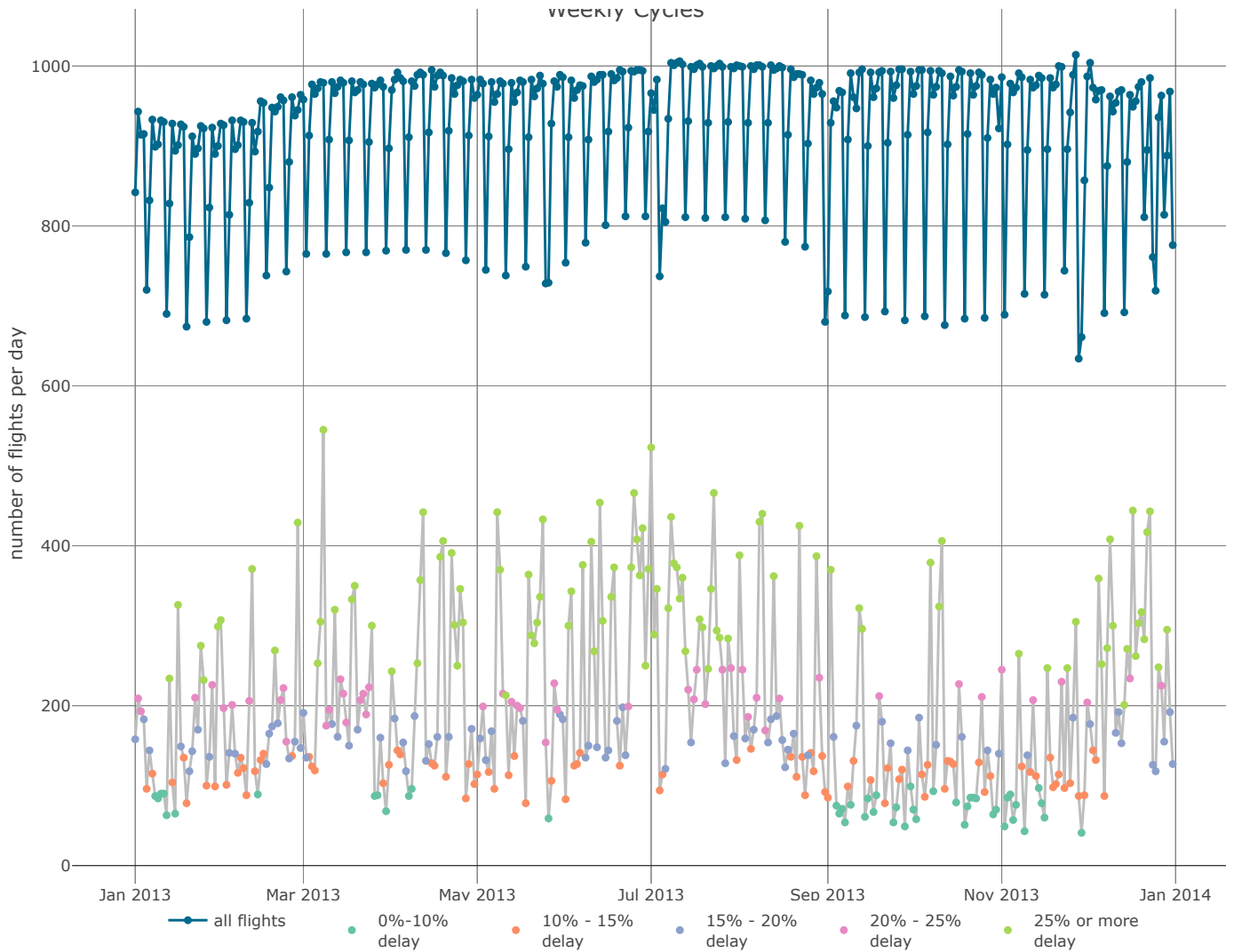
% of Flight Departures Delayed > 15 Min
LaGuardia Airport, 2013



העבר עכבר מעל הנקודות על מנת לראות פרטים נוספים

על מנת לבנות גרף זה בחרנו את הנתונים עבור שדה התעופה לה-גוארדיה בלבד, הוספנו בעבור כל טיסה משתנה בוליאני המציין האם התעכבה בהמראה יותר מ-15 דקות. בעבור כל יעד חישבנו את כמות הטיסות הכוללת ואת ממוצע הטיסות המאחרות. למידע זה חיברנו מקובץ המידע "שדות תעופה" את הנתונים הגיאוגרפיים לפי שדה היעד ושדה המוצא. הוספנו מפה שתהווה בסיס לגרף ועליה נקודות מותאמות בגודלן לכמות הטיסות לכל יעד ומחלוקות באמצעות צבעים לפי טווחי ערכי ממוצע האיחורים. כמו כן, בעבור כל יעד הוספנו מקטעים בין שדה המוצא לשדות היעד.

Temporal Effects
Weekly Cycles



העבר עכבר מעל הנקודות על מנת לראות פרטים נוספים

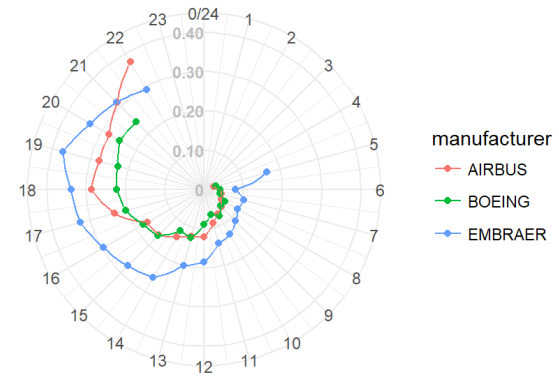
בשביל לבנות גרף זה הוספנו משתנה יום (בפורמט זמן) וחישבנו בעבור כל יום את כמות הטיסות הכוללת, את כמות הטיסות המתעכבות ואת אחוז הטיסות המתעכבות מתוך כלל הטיסות. הדפסנו את הנתונים לפי יום וכמות טיסות כוללת: כאשר את כמות הטיסות הכוללת הדפסנו בקווים ונקודות בצבע אחיד בעוד את כמות הטיסות המאחרות הדפסנו בקווים בצבע אחיד ונקודות המתחלקות לפי אחוז האיחורים בכל יום.

שאלה שלישית - ניתוח נתונים

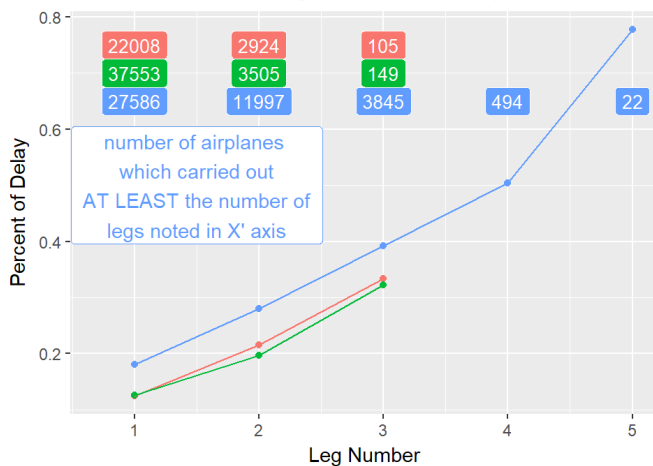
A View by Manufacturer, Airport EWR, Year 2013

	AIRBUS	BOEING	EMBRAER
all flights	25037	41207	43944
number of airplanes	651	1191	279
mean seats per plane	193.13	165.77	53.35

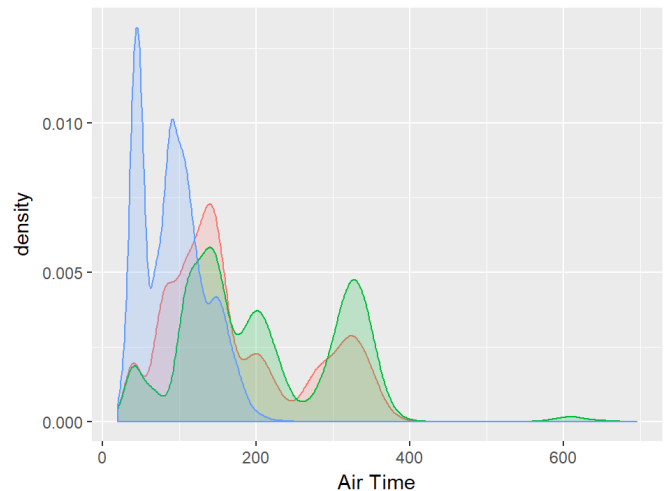
Percent of Flight Departures Delayed Throughout the Day



Percent of Flight Departures Delayed VS Leg Number for a Day



Air-Time Density



בשלב זה, ברצוננו להסביר את האיחורים בהמראות (נתיחם לאיחורים כאל כל המראה שארעה חמש עשרה דקות או יותר לאחר הזמן המתוכנן שלה) משדה התעופה "ניואר ליברטי" בניו יורק בשנת 2013. לצורך כך, נבחן את העמידה בזמנים המראה ביציאה בעבור שלוש יצרניות המטוסים הגדולות: "בואינג" (בירוק) "אמבראר" (בכחול) ו"אירבוס" (באדום), כאשר בעבור חברת "אירבוס" מדובר על נתוני "אירבוס" ו-"אירבוס תעשיות".

בגרף הראשון מוצגים על גבי מערכת קורדינטות קוטבית ממוצע האיחורים לאורך יממה בעבור כל ייצרן. כאשר הזווית מראשית הצירים עם הכיוון החיובי של ציר ה-X מייצגת את השעה. זאת בדומה לשעון המכיל 24 שעות, בעוד המרחק מראשית הצירים מייצג את ממוצע האיחורים. מגרף זה עולה תופעה בולטת. ככל שהשעה ביום מאוחרת יותר, כך עולה אחוז העיכובים בהמראות בעבור כל היצרנים. בנוסף, ניתן לשים לב שלמרות שהתופעה קיימת לכלל היצרניות, בעבור החברה "אמבראר" ניתן לראות גידול משמעותי יותר.

על מנת להבין את התופעה הנצפית החלטנו לבחון את מטוסי "אמבראר" ואופי הטיסות אל מול שאר היצרניות. תחילה הערכנו כי ייתכן שמטוסי "אמבראר" בעלי כמות נוסעים גדולה יותר ומתעכבים בשל עיכובי הנוסעים. אולם, מהתבוננות במספרים כפי שהם מופיעים בטבלה שלמעלה משמאל ניתן לראות כי ההפך הוא הנכון – המטוסים מכילים בממוצע כ-50 מושבים, בעוד "בואינג" ו"אירבוס" מכילים 3 ו-4 בהתאמה.

מהתבוננות בנתונים אלו אפשר להבחין בדבר מעניין נוסף – כמות הטיסות אשר בוצעו על ידי "אמבראר" במהלך שנת 2013 דומה לזו של "בואינג", ואף גבוהה יותר במעט, אולם כמות המטוסים ב'צ' מונה רק שישית מכמות המטוסים של "בואינג".

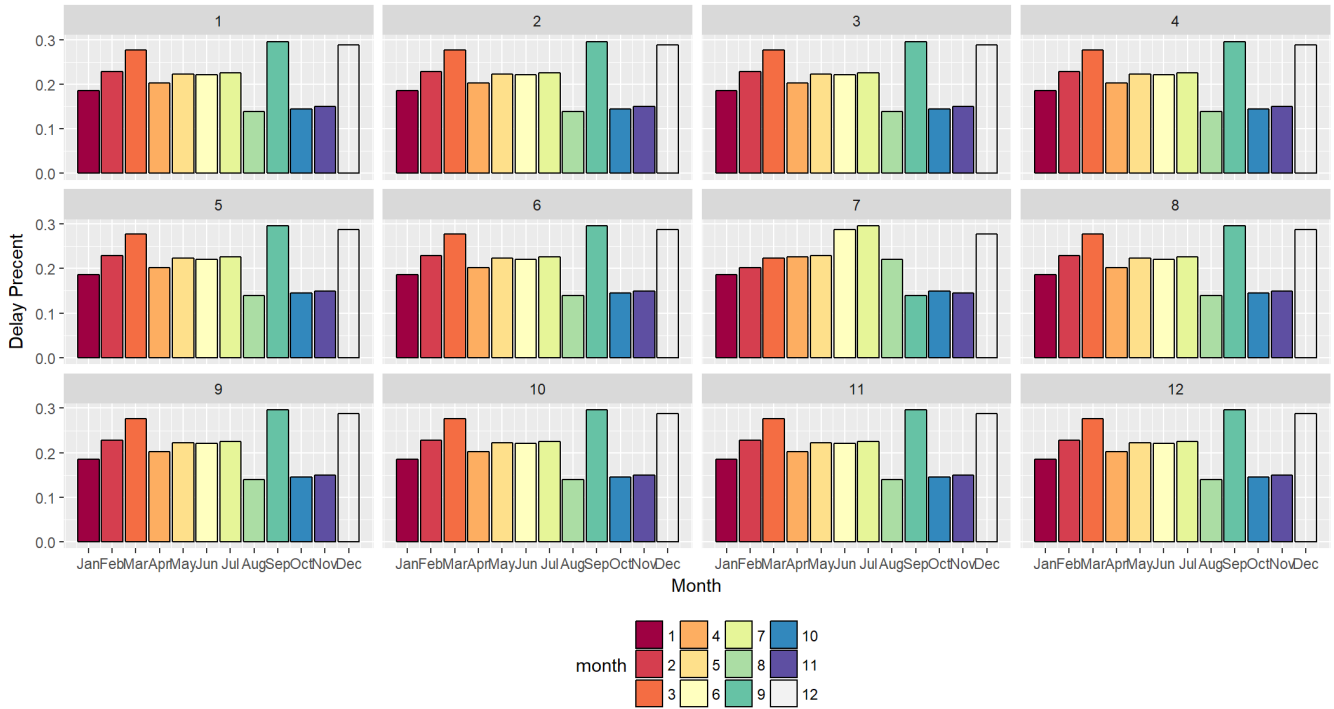
כהמשך לבחינה זאת ניתחנו את אופי הטיסות, בדגש על זמן אוויר ומרחק. בשל הדמיון הרב בנתונים בחרנו להציג רק את היסטוגרמת זמן האוויר, המופיעה בצד ימין למטה. כפי שניתן לראות, מירב מטוסי "אמבראר" מבצעים טיסות קצרות הדורשות זמן אוויר קצר יותר (ניתן לראות שני "פיקים" – סביב חמישים דקות וסביב תשעים דקות) בעוד מטוסי "בואינג" ו"אירבוס" מבצעים טיסות ארוכות בהרבה ("פיקים" ב-120, 200, 320, ובעבור בואינג ניתן לראות עלייה סביב 600 בעבור טיסות המגיעות לטווחים ארוכים במיוחד).

בשל הממצאים של היסטוגרמה זו והנתונים בטבלה עולה שאלה מהותית: כמה סבבים ("לגים" בלעז) של טיסות מבצע כל מטוס ביום? נציין שהנתונים לבדם אינם מספיקים על מנת לענות על שאלה זאת, שכן בעבור כל מטוס הממריא פעמים ביום אחד מניוארק ליברטי קיימת לפחות טיסה אחת הנוחתת בניוארק אשר אין לנו נתונים לגביה. לכן נתייחס לכמות ההמראות מניוארק ליברטי כמשנתה פרקסי אשר ייצג את כמות הסבבים למטוס ביממה.

בגרף התחתון משמאל ניתן לראות את הייצוג למספר הסבבים ביממה בציר האופקי אל מול ממוצע האיחורים (לאורך כל השנה) על הציר האנכי. בחלקו העליון של הגרף זה מופיע מספר המתאר עבור מספר הסבב למטוס את מספר הפעמים בהם אירוע זה חוזר על עצמו לאורך השנה. כך למשל מטוס שביצע ביום ארבע המראות, ייכלל בכל ארבעת המשבצות העליונות בגרף. הגרף מספק לנו מספר מסקנות חשובות: הראשונה, שכבר בעבור הסבב הראשון בכל יממה מטוסי "אמבראר" נוטים להתעכב יותר ממטוסי שאר היצרניות. השנייה, שאר החברות לא נוהגות להשתמש בשיטת סבבים. אם נסתכל על האחוזים נראה ש-87% ו-91% מהטיסות של "בואינג" ו-"אירבוס" בהתאמה מבוצעים בסבב ראשון בניגוד לרק כ-62% מהמטוסים של "אמבראר". נמשיך ונציין כי "אמבראר" מבצעת לעיתים סבבים רביעיים וחמישיים אשר מעלים באופן משמעותי את ממוצע האיחורים. אולם, נסייג טענה זאת באמירה שממוצע המטוסים שקיבלנו בעבור חמישה סבבים כנראה מושפע יותר מחריגות. שכן, כמות הטיסות הכוללת נמוכה ושוברת את קו מגמה הדומה לכל היצרניות בסבבים הראשונים.

ניתן להעלות מספר הסברים מדוע מספר סבבים רב יותר למטוס מעלה את ממוצע האיחורים. כאן, נבחר להתמקד ולסכם בדרך אחרת: בעבור כל סבב, ממוצע האיחורים עולה. כעת, ניתן להשלים את התמונה של כלל הגרפים בסיפור זה. בשל אורך הטיסות יכולה "אמבראר" לבצע כמות רבה כל כך של טיסות עם כמות נמוכה של מטוסים. בשל העיכוב בעבור כל סבב, ככל שהשעה מאוחרת יותר כך סביר יותר שמדובר בסבב מאוחר יותר ולכן אנחנו רואים לאורך היממה שממוצע האיחורים של היצרנית "אמבראר" עולה.

שאלה רביעית - בחינת הנתונים



בשאלה זו התבקשנו להוכיח שהממצאים שלנו אינם נובעים משינוי אקראי. נבחן זאת על ידי בחינה גרפית של הנתונים המקוריים ושל נתונים המוגדלים מתוך הנתונים המקוריים. נצפה לטעון כי קיימים אפקט ותבנית עונתית של ממוצע איחורים בחודש, בעוד השערת האפס הינה שממוצע זה משתנה באופן אקראי לאורך השנה והחודשים אינם תלויים אחד בשני.

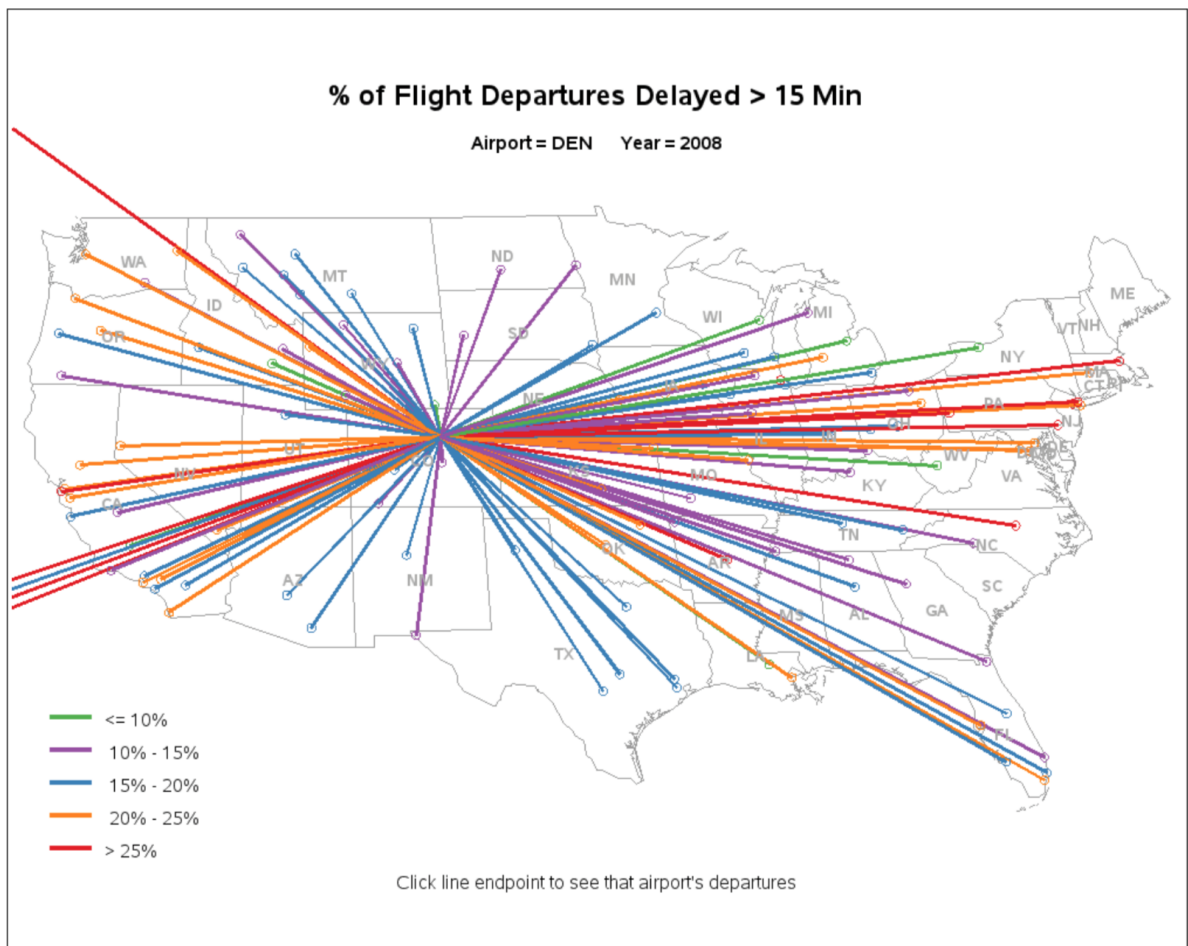
בתמונה קיימים שניים-עשר גרפים אשר מציגים את ממוצע האיחורים בכל חודש. רק באחד מן הגרפים מצוינים הנתונים האמיתיים, בעוד בשאר הגרפים קיים ערבול של הנתונים האמיתיים אשר נועד לדמות מצב אקראי בו החודשים בשנה אכן בלתי תלויים. לאחר בחינה ניתן לזהות כי גרף מספר 10 הינו הגרף האמיתי. ניתן לזהות את הגרף האמיתי על ידי צרתו שבולטת באחדות ועונתיות המראה קשר בין חלק מהחודשים הצמודים אשר לא ניתן לראות בסימולציה "אקראית".

אנו לומדים מכך שכאשר אנו מבחינים בתוצאה בעלת תבנית מסוימת עלינו לשאול את עצמנו, האם היא אקראית? בשל העובדה שאנו יכולים לזהות את הגרף הנכון מבין שאר הגרפים בעלי ההתפלגות האקראית, אנו מעריכים שהתופעה אינה אקראית ומקורה בתופעה מסוימת. במקרה המסוים הזה אנו לומדים שהשערת האפס השמרנית אינה נכונה, וכי קיימים טרנדים עונתיים. יתרה מכך, אנו לומדים על הכוח הרב של עיצוב גרפי והיכולת שלו להוות מבחן סטטיסטי. מעבר ליכולת של הגרף לספר סיפור ולהציג מידע כפי שראינו עד עתה בתרגיל; אנו למדים על היכולת שלנו לצרוב בגרף את התודעה שלנו והדעות הקדומות שאנו נושאים עמנו על הנתונים. כך, ששימוש במבחן זה עונה על השאלה "האם דעתנו על הנתונים היא ללא משוא-פנים?".

CODE

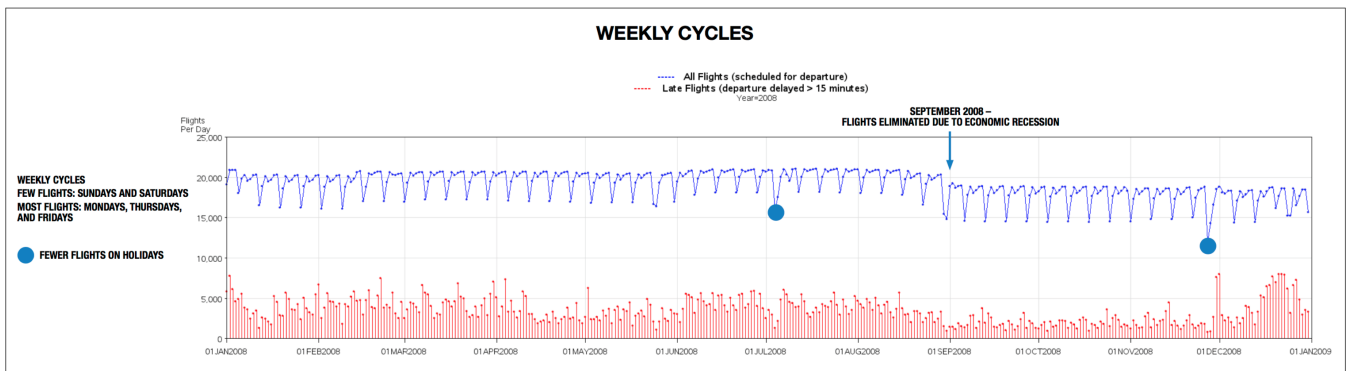
```
##<ins>שאלה ראשונה - ניתוח גרפים קיימים</ins>
```

```
library(magick)
frink <- image_read("https://i.imgur.com/l5eR5HH.png")
image_trim(frink)
```

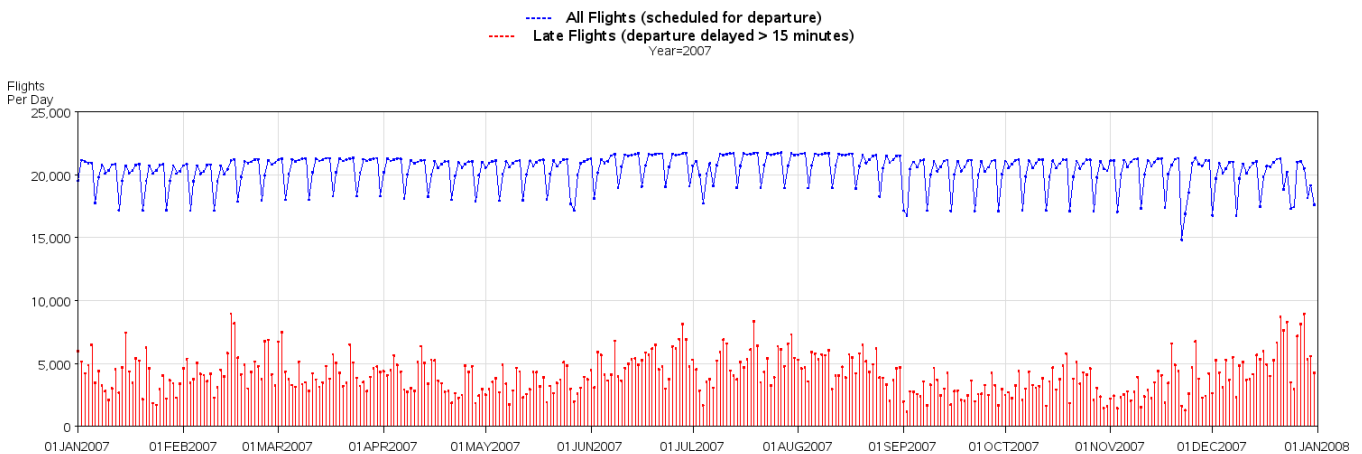


```
library(magick)
IM <- image_read("https://image.ibb.co/hmLnRx/Flight_Figure_2.png")
image_trim(IM)
```

TEMPORAL EFFECTS



```
library(magick)
IM <- image_read("http://robslink.com/SAS/democd40/asa00120.gif")
image_trim(IM)
```



##Q2

#Librarys

library("nycflights13")

library(plotly)

library(plyr)

flights <- flights

flights\$large_delay <- flights\$dep_delay>15

EWR <- subset.data.frame(x=flights,subset = (origin=="EWR"))

JFK <- subset.data.frame(x=flights,subset = (origin=="JFK"))

LGA <- subset.data.frame(x=flights,subset = (origin=="LGA"))

summarize_by_airport <- ddply(LGA,~dest,summarise,all_flights=length(year),delay_prob=mean(large_delay,na.rm = TRUE))

subset_airports <- airports[which(airports\$faa %in% summarize_by_airport\$dest),]

names(subset_airports)[names(subset_airports)=="faa"] <- "dest"

summarize_by_airport <- merge.data.frame(x = summarize_by_airport,y =subset_airports)

summarize_by_airport <- select(summarize_by_airport,dest,all_flights,delay_prob,name,lat,lon)

summarize_by_airport\$ori_lat <- airports\$lat[airports\$faa=="LGA"]

summarize_by_airport\$ori_lon <- airports\$lon[airports\$faa=="LGA"]

summarize_by_airport\$groups <- cut(summarize_by_airport\$delay_prob, breaks=c(-Inf,0.10,0.15,0.20,0.25,1),c("0%-10% ", "10% - 15%", "15% - 20%", "20% - 25%", "25% or more"))

geo <- list(

scope = 'usa',

projection = list(type = 'albers usa'),

showland = TRUE,

landcolor = toRGB("gray95"),

countrycolor = toRGB("gray80")

)

plot_geo(locationmode = 'USA-states',color = ~groups, colors = c("forestgreen","darkorchid","dodgerblue3","darkorange","firebrick1")) %>%

add_markers(

data = summarize_by_airport, x = ~lon, y = ~lat, text = ~name,

size = ~all_flights, alpha = 1, showlegend = TRUE

)%>%

add_segments(

#data = group_by(f, groups),

x = ~ori_lon, xend = ~lon,

y = ~ori_lat, yend = ~lat,

alpha = 0.55, size = I(1), showlegend = FALSE

)%>%

layout(

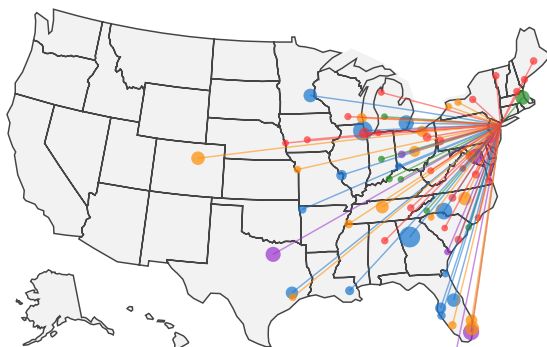
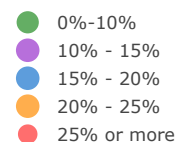
#Legend=List()

title = '% of Flight Departures Delayed > 15 Min
 LaGuardia Airport, 2013',

geo = geo, showlegend = TRUE#, height=350

)

% of Flight Departures Delayed > 15 Min
LaGuardia Airport, 2013

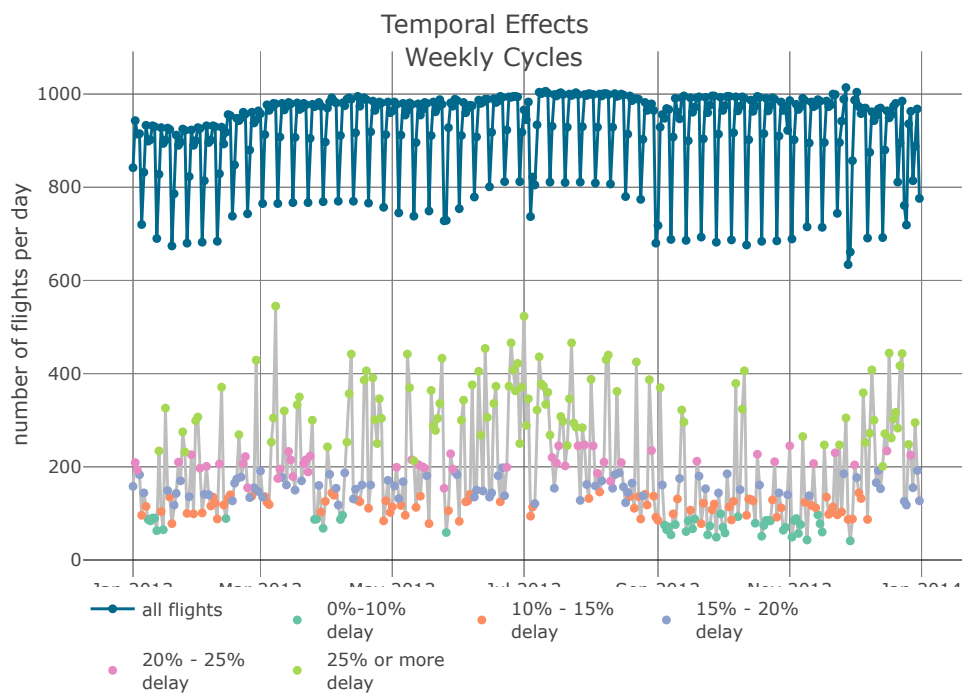



```

flights$day <- as.Date(flights$time_hour)
summarize_by_day <- ddply(flights,~day,summarise,all_flight_per_day=length(year),delay=sum(large_delay,na.rm = TRUE))
summarize_by_day$prop <- summarize_by_day$delay/summarize_by_day$all_flight_per_day
summarize_by_day$groups <- cut(summarize_by_day$prop, breaks=c(-Inf,0.10,0.15,0.20,0.25,1),c("0%-10%", "10% - 15%", "15% - 20%", "20% - 25%", "25% or more"))

plot_ly(summarize_by_day, x = ~day, y = ~all_flight_per_day, name = 'all flights', type = 'scatter', mode="lines+markers", color=I("deepskyblue4")) %>%
  layout(legend = list(x = 0.5, y = -850,orientation = "h",xanchor = "center")) %>%
  add_trace(y = ~delay, name = 'delay', mode="line",color=I("grey"), showlegend = FALSE) %>%
  add_trace(y = ~delay, name = 'delay',color=~groups, mode="lines+markers", colors = c("forestgreen","darkorchid","dodgerblue3","darkorange","firebrick1"))%>%
  layout(title = 'Temporal Effects <br> Weekly Cycles', xaxis=list(title=" "),yaxis=list(title="number of flights per day"))

```




```
##Q3

rm(list=ls())

library("ggplot2")
library("nycflights13")
library("plyr")
library("plotly")

flights <- flights
weather <- weather
planes <- planes
airports <- airports
airlines <- airlines

flights$large_dep_delay <- flights$dep_delay>30
flights$large_arr_delay <- flights$arr_delay>30
flights$Eday <- as.Date(flights$time_hour)

EWR <- subset.data.frame(x=flights,subset = (origin=="EWR"))

EWR <- EWR[order(EWR$tailnum),]
planes <- planes[order(planes$tailnum),]

EWR <- merge(x=EWR,y=planes, by.x = 'tailnum',by.y = 'tailnum',all.x = TRUE)

library("dbplyr")
library("lubridate")

man <- subset(EWR,EWR$manufacturer %in% c("BOEING","AIRBUS INDUSTRIE","AIRBUS","EMBRAER"))

man$manufacturer <- ifelse(man$manufacturer=="AIRBUS INDUSTRIE","AIRBUS", man$manufacturer)

A <- ddply(man,.(manufacturer,hour),summarise,delay=sum(large_dep_delay,na.rm = T),all=length(day))
A$precent=A$delay/A$all

man_by_houer <- subset(A,(hour>4)&hour<23)

man_by_houer_plot <- ggplot(man_by_houer,aes(x=hour,y=precent,color=manufacturer))+
  ggtitle("Percent of Flight Departures Delayed \n Throughout the Day")+
  geom_point()+
  theme_minimal()+
  geom_line()+
  scale_x_continuous(breaks=c(0:24), limits=c(0, 24))+
  annotate("text", x = 0, y = seq(0,0.4,by=0.1), label = c("0", "0.10", "0.20", "0.30","0.40") , color="grey", size=3 , angl
e=0, fontface="bold", hjust=1) +
  ylim(0,0.4)+
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position="right",
    axis.title = element_blank(),
    axis.text.y=element_blank()#,
    # plot.margin = unit(rep(-2,4), "cm")
  ) +
  coord_polar(start = 0)

man_by_air_time <- ggplot(man,aes(x=air_time,color=manufacturer,fill=manufacturer))+
  ggtitle("Air-Time Density ") +
  geom_density(alpha=0.2)+
  theme(legend.position="none",
    plot.title = element_text(hjust = 0.5))+
  xlab("Air Time")

man <- man[order(man$tailnum,man$time_hour),]

man$order <- NA

TAIL <- man$tailnum[1]
DAY <- man$Eday[1]
i <- 0

for (k in 1:nrow(man)) {
  if (man$tailnum[k]==TAIL){
    if (man$Eday[k]==DAY){
      i <- i+1
    }
  }
}
```

```

    } else{
      DAY <- man$Eday[k]
      i <- 1
    }
  } else{
    TAIL <- man$tailnum[k]
    i <- 1
  }
  man$order[k] <- i
}

library(ggExtra)

data <- ddply(man,.(manufacturer,order),summarise,all=length(day),mean=mean(large_dep_delay,na.rm = T))

data <- data[order(data$order,data$manufacturer),]

library(ggrepel)

by_lag <- ggplot(data,aes(x=order,y=mean,color=manufacturer))+
  geom_point()+
  ggtitle("Percent of Flight Departures Delayed VS Leg Number for a Day")+
  xlab("Leg Number")+
  ylab("Percent of Delay")+
  geom_line()+
  geom_label(aes(label = all,
                 fill = factor(manufacturer)), color = 'white',
             size = 3.5, y=c(0.75,0.7,0.65,0.75,0.7,0.65,0.75,0.7,0.65,0.65,0.65))+
  theme(legend.position="none",
        plot.title = element_text(hjust = 0.5))+
  geom_label(aes(label = c("number of airplanes which carried out \n AT LEAST the number of \n legs noted in X' axes "), y=c
(0.5), x=c(2)))

TD <- ddply(man,~manufacturer,summarise,all_flights=length(day),tail_number=length(unique(tailnum)),seats=round(mean(seats),
2))
library(data.table)

library(ggpubr)
t_TD <- transpose(TD[,-1])
colnames(t_TD) <- TD[,1]

tab <- ggtexttable(t_TD, rows = c("all flights","number of airplanes","mean seats per plane"),
  theme = ttheme(
    colnames.style = colnames_style(color = "white", fill = c(rgb(red = 248/256,green = 118/256,blue = 109/256),
      rgb(red = 0/256,green = 186/256,blue = 56/256),
      rgb(red = 97/256,green = 156/256,blue = 255/256)))
  )
)

tab <- table_cell_bg(tab, row = 2, column = 2, linewidth = 1,
  fill=rgb(red = 248/256,green = 118/256,blue = 109/256,alpha = 0.65), color = "white")
tab <- table_cell_bg(tab, row = 3, column = 2, linewidth = 1,
  fill=rgb(red = 248/256,green = 118/256,blue = 109/256,alpha = 0.65), color = "white")
tab <- table_cell_bg(tab, row = 4, column = 2, linewidth = 1,
  fill=rgb(red = 248/256,green = 118/256,blue = 109/256,alpha = 0.65), color = "white")

tab <- table_cell_bg(tab, row = 2, column = 3, linewidth = 1,
  fill=rgb(red = 0/256,green = 186/256,blue = 56/256,alpha = 0.65), color = "white")
tab <- table_cell_bg(tab, row = 3, column = 3, linewidth = 1,
  fill=rgb(red = 0/256,green = 186/256,blue = 56/256,alpha = 0.65), color = "white")
tab <- table_cell_bg(tab, row = 4, column = 3, linewidth = 1,
  fill=rgb(red = 0/256,green = 186/256,blue = 56/256,alpha = 0.65), color = "white")

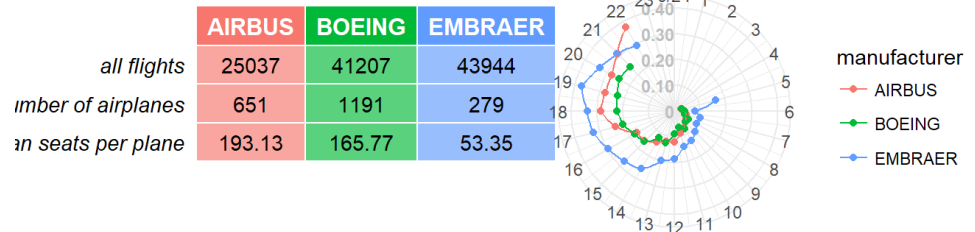
tab <- table_cell_bg(tab, row = 2, column = 4, linewidth = 1,
  fill=rgb(red = 97/256,green = 156/256,blue = 255/256,alpha = 0.65), color = "white")
tab <- table_cell_bg(tab, row = 3, column = 4, linewidth = 1,
  fill=rgb(red = 97/256,green = 156/256,blue = 255/256,alpha = 0.65), color = "white")
tab <- table_cell_bg(tab, row = 4, column = 4, linewidth = 1,
  fill=rgb(red = 97/256,green = 156/256,blue = 255/256,alpha = 0.65), color = "white")

library(gridExtra)

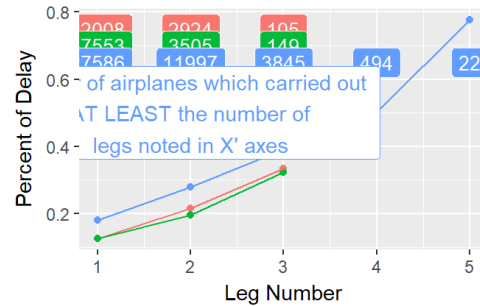
p1 <- ggarrange(tab,man_by_houer_plot)
p2 <- ggarrange(by_lag,man_by_air_time)
ggarrange(p1,p2,ncol=1,nrow=2,labels = "A View by Manufacturer, Airport EWR, Year 2013")

```

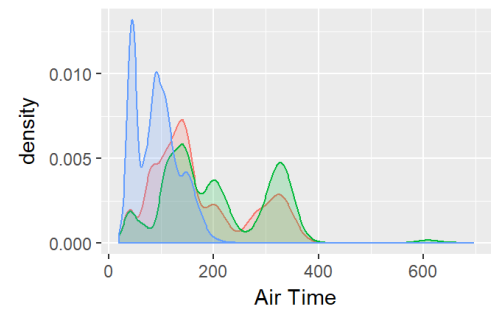
A View by Manufacturer of Flight Departure Delay Throughout the Day



t of Flight Departures Delayed VS Leg Numl



Air-Time Density



```

flights$large_dep_delay <- as.integer(flights$dep_delay>15)
flights$large_arr_delay <- flights$arr_delay>15
flights$Eday <- as.Date(flights$time_hour)

delay_by_month <- ddply(flights,~month,summarise,big_delay=mean(large_dep_delay,na.rm = T))

data_for_test <- cbind(sample_num=rep(1:12,each=12),month=rep(1:12,times=12),big_delay=rep(sample(x=delay_by_month$big_delay,size=12),12))
data_for_test <- as.data.frame(data_for_test)

data_for_test[which(data_for_test$sample_num==7),-1] <-delay_by_month
data_for_test$month <- as.factor(data_for_test$month)
month_name <- month.abb[c(1:12)]

ggplot(data_for_test, aes(x=as.integer(month),y=(big_delay),fill=month))+
  theme(legend.position="bottom")+
  guides(colour = guide_legend(nrow = 2,ncol=6))+
  geom_bar(stat="identity",color="black")+
  scale_x_continuous(breaks =1:12, labels = month_name)+
  ylab("Delay Percent")+
  xlab("Month")+
  facet_wrap(~sample_num)+
  # theme_minimal(base_size = 11, base_family = "")+
  scale_fill_brewer(palette="Spectral")

```

