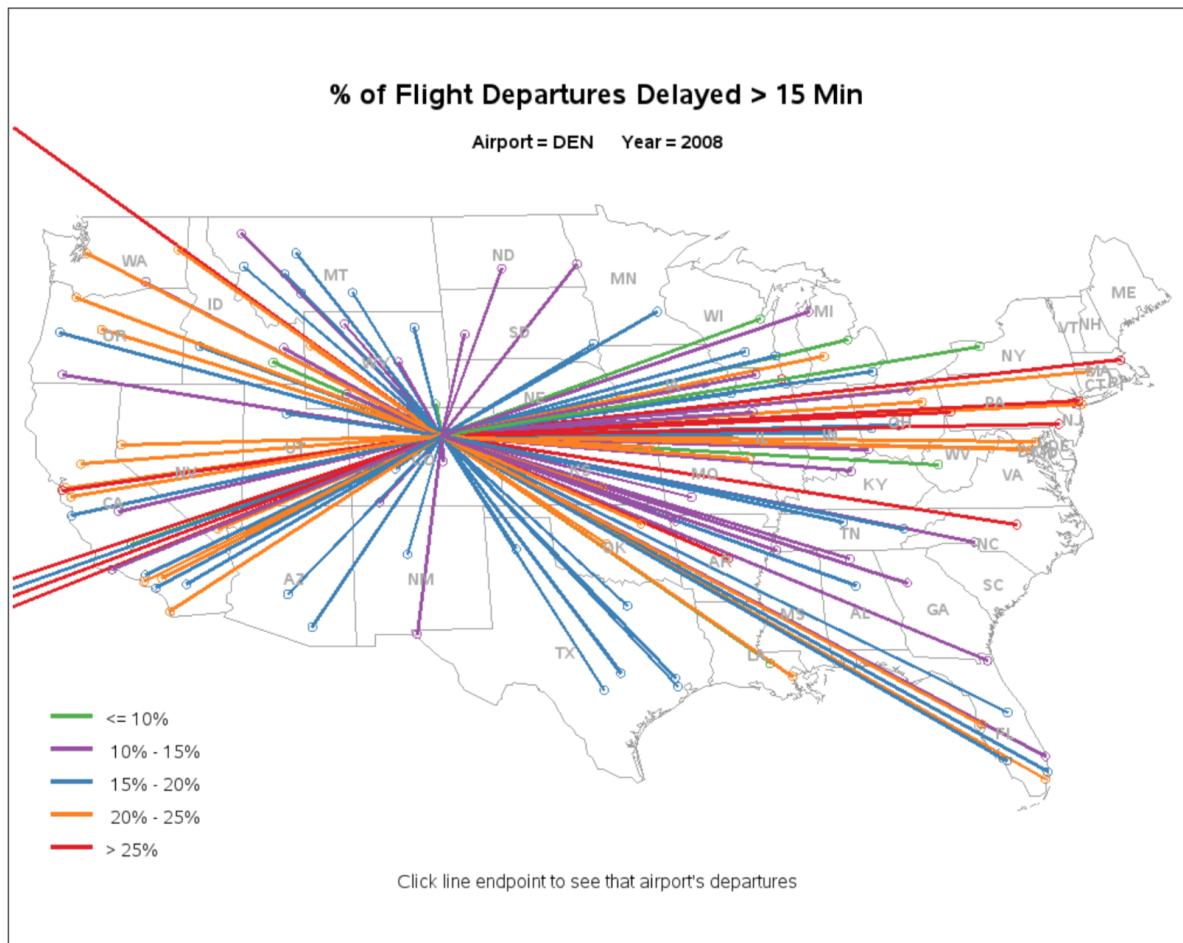


מעבדה ראשונה - ניתוח נתונים טיסות

עד זיו וספיר חנצין

שאלה ראשונה - ניתוח גרפים קיימים

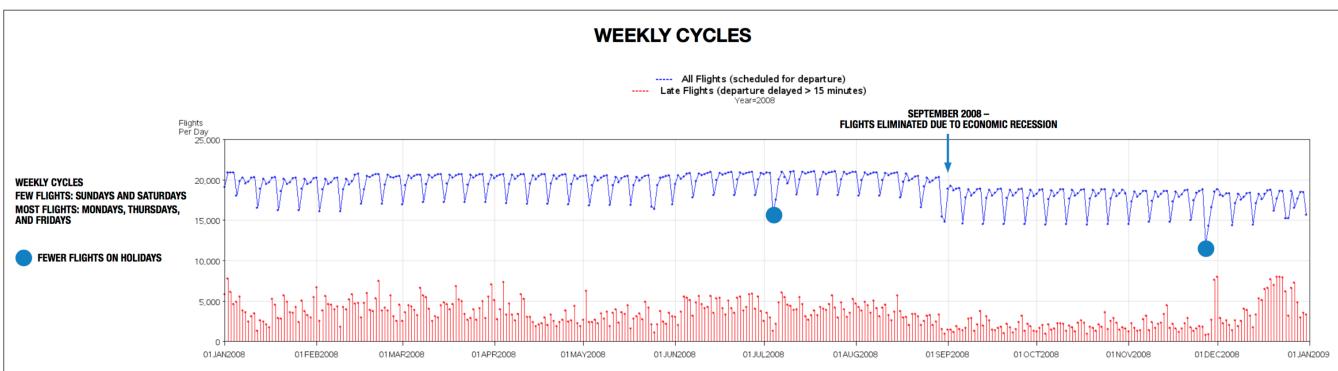


הגרף מציג את ספורה של נמל התעופה דנוור, בקולורדו, בשנת 2008. את הספורה שמספר הגורף ניתן לחלק לשני רבדים: ברובד הפשטוט ביותר מוגנתת תנועה של כל היחידים שיוצאים מדןור למדינות ארצות הברית. כאשר, בעבר כל יעד, מזג לאזפה אחוז הטיסות המתעכבות בỜמה ערך ממוצע (מעל 15 דקות). ברובד המרכיב יותר והנתון לפשרות, ניתן לראות תמונה המספרת כי "הסתברות" לעיכוב משמעותי גדולה יותר בעבר יעדים רחוקים יותר (כך לדוגמה ההסתברות לעיכוב בטיסות לאלאסקה, הוואי, הנולו, או החוף המזרחי, דהו' משמעותית מהסתברות לעיכוב בominatorה, ויימנג, ניו מקייקן, טקסס ועוד).

לעתינו, ישנו כישלון העולה מן הגרף בלספר סיפור מרכיב יותר. בעקבות היכילון והתוואה היצירתיות עלות מספר שאלות נוספת: הריאונה, שעולה כמעט באופן מייד: מהן כמהות הטיסות לכל יעד? שאליה זו פותחת דלת לסתוגיה האם קיים קשר בין העיכובים? שאליה נספה שאליה ותעדו למקד את הסיפור שאותו הגרף מספר מהירות הטיסות אל שדה התעופה דנוור או רק מתוtal שדה התעופה דנוור אל מול שדות תעופה אחרים? האם ש זו טיסות רוכות יותר וכן עיכובים הרבה יותר? האם ניתן לראות דפוסים דומים של ההסתברות בגובה יותר לעיכובים גם בשדות תעופה אחרים? פתרון שאליה זו שבא לידי ביטוי בהוספת גרפים דומים של נמלי תעופה אחרים, מעיד על יכולתו של הגרף הנוכחי לספר סיפור עמוק ומשמעותי.

נציע לספר את הסיפור בצהורה קצת שונה: ברמה הגרפית, קווים דקים יותר, ויציג אחוז הטיסות המתעכבות באמצעות צבע הנקודה במקומם, "יצוג מצומצם" הפורק את הצפיה בגרף לנעימה ובזרה יותר. ברמה הספורית, נציג להציג באמצעות גודל הנקודות את כמהות הטיסות היוצאות ליעד, זאת כדי לספק עוד מידע לצופה. כמו כן, כדי לענות על השאלה – האם האמם ההסתברות לעיכוב יעדים רחוקים יותר? נציג להוסיף על הגרף טבעות ולהציג את ההסתברות לעיכוב משמעותי עבורי כל טבעת. ניתן גם להציג מספר גרפים במקביל של שדות תעופה שונים.

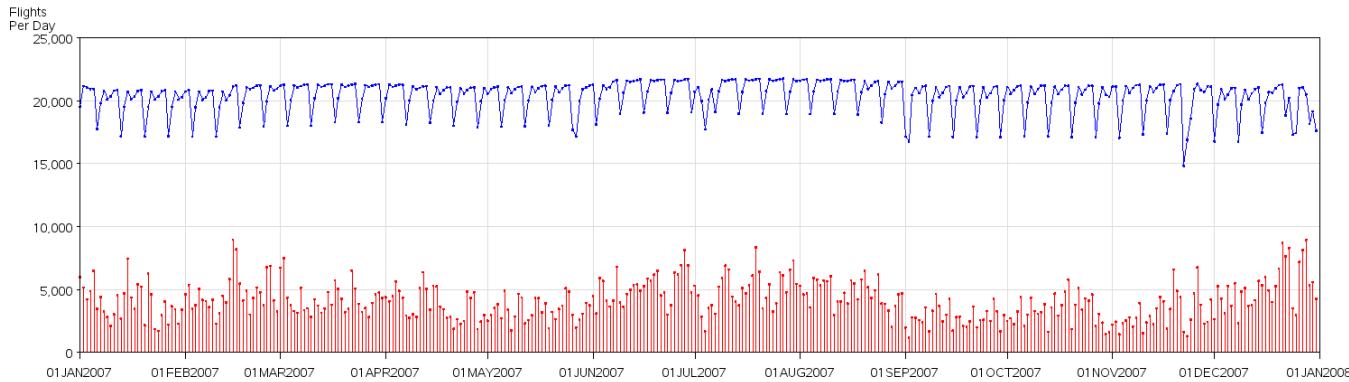
TEMPORAL EFFECTS



בGBT ריאשוני על הגראף ניתן לראות שסתור העיקרי שמסופר הוא כמהות האיתורים ביום, במשך שנת 2008. את הסיפור של הגראף ניתן לבדוק בשני רבדים, שבועי ושנתי. בrama השבועית ניתן כי בכל שבוע ביום ראשון ושבת מתוכנות פחתה טיסות מbasar השבוע. כמו כן ניתן לראות, בrama חקלית בלבד, כי ביום אלו כמהות האיתורים עולה. נציין כי מסקנה זו אינה חד משמעית לאורור הגראף. בrama השנתית מנסה הגראף לספר שבסהilar החגים כמהות הטיסות יורדת דמיית ועולה מיד לאחר החג בחזרה. כמו גם, שבקבוקת המשבר הכללי בשנת 2008 ירדו כמהות הטיסות בזרחה משמעותית – נתיחס לכך זהה זאת בהמשך כאשר נציג שיפורים. לדעטנו, צורת המuna לשאלות הללו היא חקלית בלבד ואין מספקת. ככל העקריים של הגראף נובעים מכך שכrama היומי והשנתית קשה לדודת לפרטים הקטנים הן בשל כמהות הנזומות, צורת הגאנטם ופריטים. ככל זה בולט במיוחד במקרים הטיסות מתוארכות בזרחה רציפה בעוד כמהות הטיסות המתעכבות מתוארכות בזרחה.

ברובד השנתי מתגלה ככל עיקרי שבא לידי ביטוי בחומר התיכון לאירועים משמעותיים הפועלים במהלך השנה. כך למשל מציניעים על הגראף רק חלק מהחגים. לדוגמה, סיבוב חג המולד אוليل הקדושים לא ניתן לראות שניי דומה. הגאה זו מעודדת את התמונה ואורמת להסקת מסקנות שאנו נכוון בהכרח. כאמור, מצין בגרף תחילת המשבר הכללי בספטמבר 2008 בה ירדו כמהות הטיסות בצורה משמעותית, הסבר נוסף שנitin היה להעלות על הדעת הוא פטיחת שנת הלימודים. כפי שmobius, מבט על הגראף בשנים האחרונות מראה לנו שההטעה חוזרת על עצמה ושההסבר המובא כ"חודי לשנת 2008 שגוי".

All Flights (scheduled for departure)
Late Flights (departure delayed > 15 minutes)
Year=2008



כשל נספ הינו חוסר אינפורמציה: האם המידע המתואר מייצג את כל ארצות הברית או ממל עופה מסוים? האם הוא כולל טיסות פנים וחוץ אחד?

על מנת לשפר את הגראף נציג מספר נקודות עיקריות: הראשונה בהם, תצוגה וריפה יותר, תוך הוספה נוספת גוף של צבע כדי ליצג באמצעותו את כמהות הטיסות המתעכבות. בתוצאות הנוכחית של הגראף לא ניתן להשוות בזרחה טוביה בין העיכובים בחודשים בהן ישן טיסות מסוימות יותר או הרבה.

כמו כן, במידה והטעה היא שסביר חגים יש ירידת בכמות הטיסות הכלולות, יש להציג את כל החגים. על מנת להשלים את התמונה בזרחה הרחבה ביוטר מומלץ להראות גם שיטים נוספים ולסמן אירועים מחזוריים לרלוונטיים נוספים.

שאלה שנייה - שחזור גרפים

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

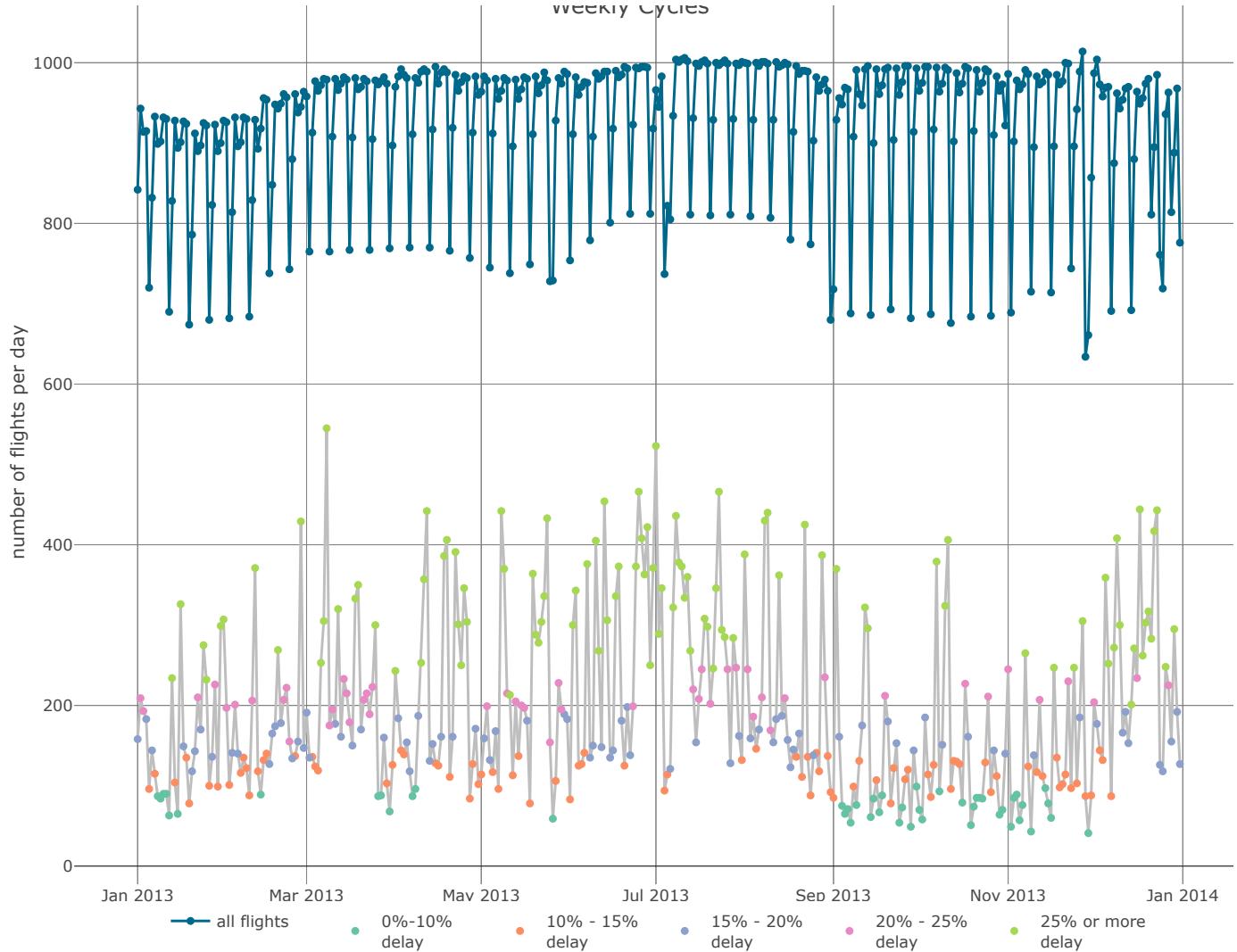
- 0%-10%
- 10% - 15%
- 15% - 20%
- 20% - 25%
- 25% or more



העבר עבר מעל הנקודות על מנת לראות פרטיים נוספים

על מנת לבנות גראף זה בחרנו את הנתונים עבור שדה התעופה לה-גוארדיה בלבד, הוספנו בעבר כל טיסה משתנה בולאייני המציין אם התעכבה בהמטרה יותר מ-15 דקות. עבור כל יעד חישבנו את כמהות הטיסות הכלולות ואת ממוצע הטיסות המאוחרות. למידע זה חיברנו מוקבץ המידע "שדות תעופה" את הנתונים הגיאוגרפיים לפי שדה היעד שדה המוצא. הוספנו מפה שתהווה בסיס לוגרף ועליה נקודות מותאמות בגודל לכמות הטיסות לכל יעד ומחלקות באמצעות צבעים לפי טווחי ערכי ממוצע האיתורים. כמו כן, בעבר כל יעד הוספנו מקטעים בין שדה המוצא לשדות היעד.

Temporal Effects Weekly Cycles



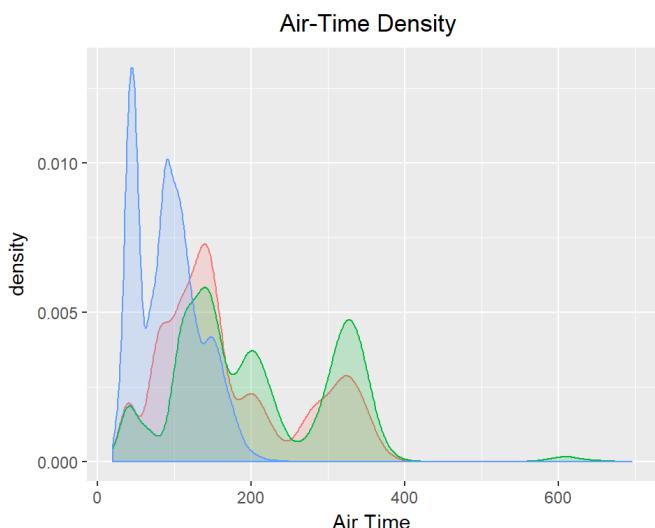
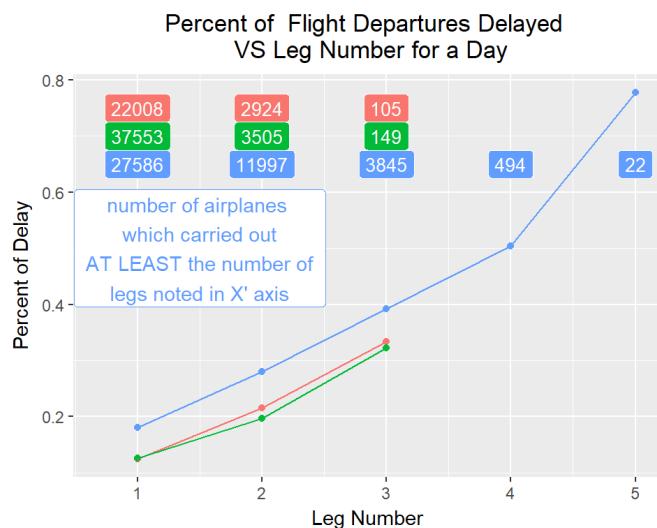
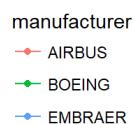
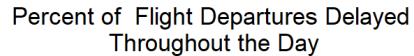
העבר עכבר מעל הנקודות על מנת לראות פרטיים נוספים

בשביל לבנות גרפ' זה הוספנו משתנה יומ (בפורמט זמן) וחישבנו בעבר כל ים את כמות הטיסות הכוללת, את כמות הטיסות המתעכבות ואת אחוז הטיסות המתעכבות מתוך כלל הטיסות. הדפסנו את הנתונים ליום וכמות טיסות כוללת: כאשר את כמות הטיסות הכוללת הדפסנו בקווים ונקודות ביצב אחד בעוד את כמות הטיסות המאחרות הדפסנו בקווים ביצב אחד ונקודות המתחלקות לפי אחוז האחורים בכל יום.

שאלה שלישית - ניתוח נתונים

A View by Manufacturer, Airport EWR, Year 2013

| | AIRBUS | BOEING | EMBRAER |
|-----------------------------|--------|--------|---------|
| <i>all flights</i> | 25037 | 41207 | 43944 |
| <i>number of airplanes</i> | 651 | 1191 | 279 |
| <i>mean seats per plane</i> | 193.13 | 165.77 | 53.35 |



בשלב זה, בראכינו להסביר את האיחורים בהמראות (וניחס לאיחורים כל המראה שארעה לאחר הזמן המתוכנן שלה) מושך התעופה "ניאור ליברטי" בניו יורק בשנת 2013. לזכור כך, נבחן את העמידה בזמן המראה ביציאה בעבר שלוש יצירויות המתוים הגדולות: "ボイング" (בירוק) "אמבראר" (כחול) ו"איירבוס" (באדום), כאשר בעבר חברת "איירובס" מדווח על נתוני "איירבוס"-ן "איירבוס תעשיית".

בגרף הראשון מוצגים על גבי מערכת קורדיוטים ממוצע האיתורים לאורק יממה בעבר כל יציר. כאשר חזויי מראות הצירים עם הכוונן החיווי של ציר ה-X-מייצג את השעה. זאת בדומה לשעון המכיל 24 שעות, בעוד המרחק מראות הצירים מייצג את ממוצע האיתורים. מגזר זה עולה焘עה בולטות. ככל שהשעה ביום מוארכת יותר, כך עולה אחוז העיכובים בהמראות בעבר כל היצירויות. בנוסף, ניתן לשים לב שלמרות שהתופעה קיימת לכל היצירויות, בעבר החברה "אמבראר" ניתן לראות גידול משמעותי יותר.

מהתבוננות בתנוטות אלו אפשר לבחון בדבר מעניין נוסף – כמהו הטיסות אשר בוצעו על ידי "אמבראר" במהלך שנת 2013 דומה לו של "בוינג", אף גבוה יותר במעט, על מנת להבהיר את התופעה הנכפית החלהנו לבחון את מטוסי "אמבראר" ואופיו הטיסות אל מול שאר היצרניות. תחילה הערכנו כי יתכן שמטוסי "אמבראר" בעלי כמה נזעים גדולה יותר ומטעכנים בשל עיקורי הנזעים. אולם, מהתבוננות במספרים כפי שהם מופיעים בטבלה שלמעלה משמאלי ניתן לראות כי ההפרש הוא הנכון – המטוסים מכילים בממוצע 50 מושבים, בעוד "בוינג" ואירובוס מכילים פי 3-4 בהתאם.

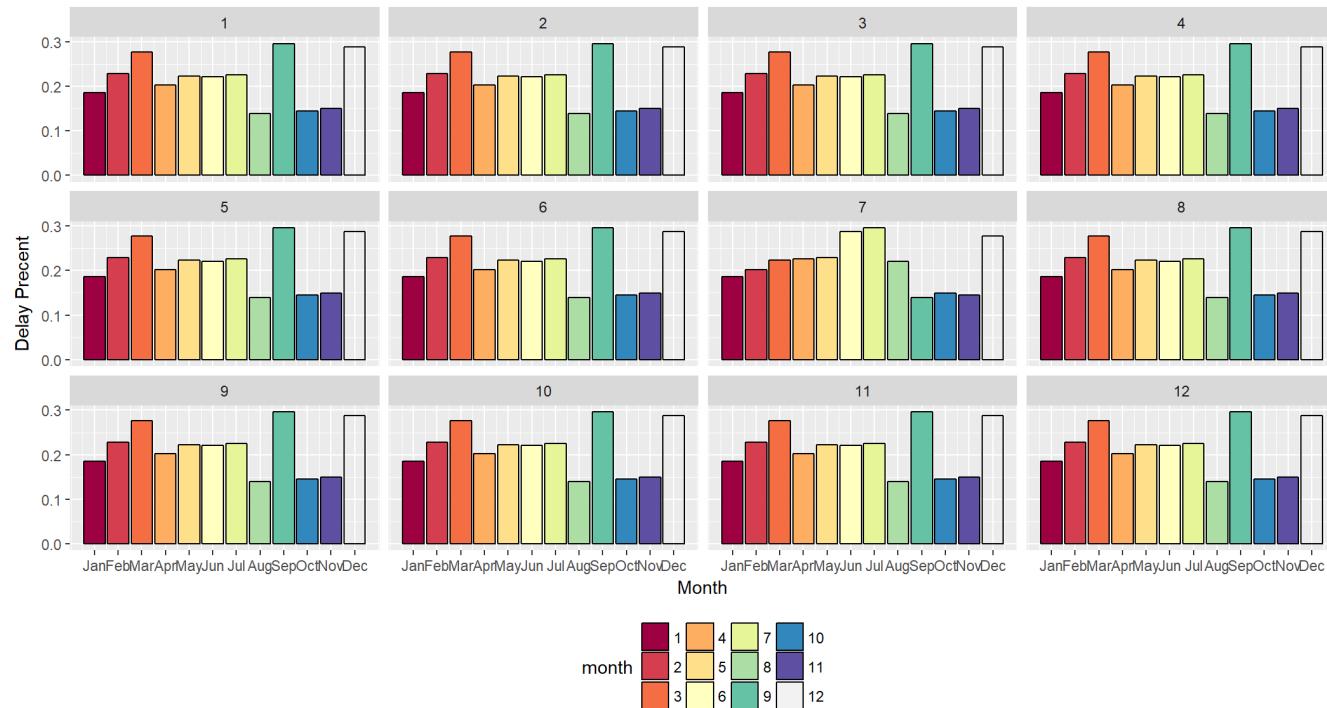
כמובן לבחינה זאת ניתחנו את אופי הטיסות, בדגש על זמן אויר ומרחק. בשל הדמיון הרב בנתונים בחורנו להציג רק את היחסות זמן האויר, המופיעה בכך ימ' בלבד. כפי שניתן לראות, מירב מטוסי "אמבראר" מבצעים טיסות קצרות בזמן קצר יותר (ניתן לראות שני "פיקט" – סביר חמישים דקות וסביר תשעים דקות) בעוד

בשל הממצאים של היסטוגרמה זו והנתונים בטבלהعلاה שאללה מהותית: כמה סבבים ("לגים" בלבד) של טיסות מבצע כל מטוס ביום? נציין שהנתונים לבדים אינם מסתיקים על מנת לענות על שאללה זאת, שכן בעבר כל מטוס הממרא פעם ביום אחד מニアרק ליברטי' קיימת לפחות הנוחותה בニアרק אשר אין לנו נתונים לגבייה. لكن נתיחס לכמות ההמראות מニアרק ליברטי' כמשתנה פרוקט' יציג את כמות הסבבים למטוס ביוםמה.

בגרף התחתון משמאל ניתן לראות את היצוג למספר הסבבים ביממה בציר האופקי אל מול ממוצע האיחורים (לאורך כל השנה) על הציר האנכי. בחלקו העליון של הגרף זה מופיע מספר המתאר עבור מטוס אחד מספר הפעמים בהם אירוע זה חורף במהלך השנה. קר' למשל מטוס שביצע ביום ארבע מראות, "ילך בכל ארבעת המשבצות העליונות בגראף. הגרף מופיע מסקנות חשובות: הראשונה, שכבר בעבר הסבב הראשון בכל יממה מטוסי "אמבראר" נוטים להתקע יותר ממוטסיו' שארכ' היצירוניות. השניה, שגם החברות לא נהוגות להשתמש בשיטות סבבים. אם נסתכל על האחדים נראה ש- 87%- 91% מהטיסות של "בוינג"-ו-"אראבוס"

בהתאם מוצעים בסבב ראשון ביאיג לרק כ- 62% מהמטוסים של "אמבראר". נזכיר ונציין כי "אמבראר" מוצעת ליעטים סבבים רביעים וחמשיעים אשר מעלים באופן שמעוני את ממוצע האיחורים. אולם, נזכיר טענה זאת באמרה שמדובר המטוסים שקיבלו בעבר חמשה סבבים כנראה מושפע יותר מחיריגות. שכן, כאמור הטיסות הכלולות וכן אשורות את גובה הרים לכל היבטים הכרחיים הראשוניים

ניתן להעלות מספר סברים רב יותר לטעות מסוימת האיתוריות. אכן, נבחר להתמקדש ולסכם בדרך אחרת: בעבר כל סבב, ממוצע האיתוריות עולגה. בעוד, ניתן להשלים את התמונה של כלל הגאים במסיפור זה. בשל אורך הטיסות יכולת "אמבראר" לבצע כמה רבה כל כך של טיסות עם כמות נמוכה של מטוסים. בשל העיכוב בעבר כל סבב, ככל שהשעה מאוחרת יותר אף סביר יותר שמדובר בסבב מואחר יותר וכןikananchorn רואים לאחר מכן ממוצע האיתוריות של היצרנית "אמבראר".

שאלה רביעית - בחינת הנתונים

בשאלה זו התקשנו להוכיח שהממצאים שלם אינם נובעים ממשוני אקרה. בבחן זאת על ידי בחינה גרפית של הנתונים המקוריים ושל נתונים המוגבלים מתוך הנתונים המקוריים. נרצה לטעון כי קיימים אפקט ומבנה עונתי של ממוצע איחורים בחודש, בעוד השערת האפס הינה שמדובר עזזה משתנה באופן אקרה לאורך השנה והחודשים אינם תלוים אחד בשני.

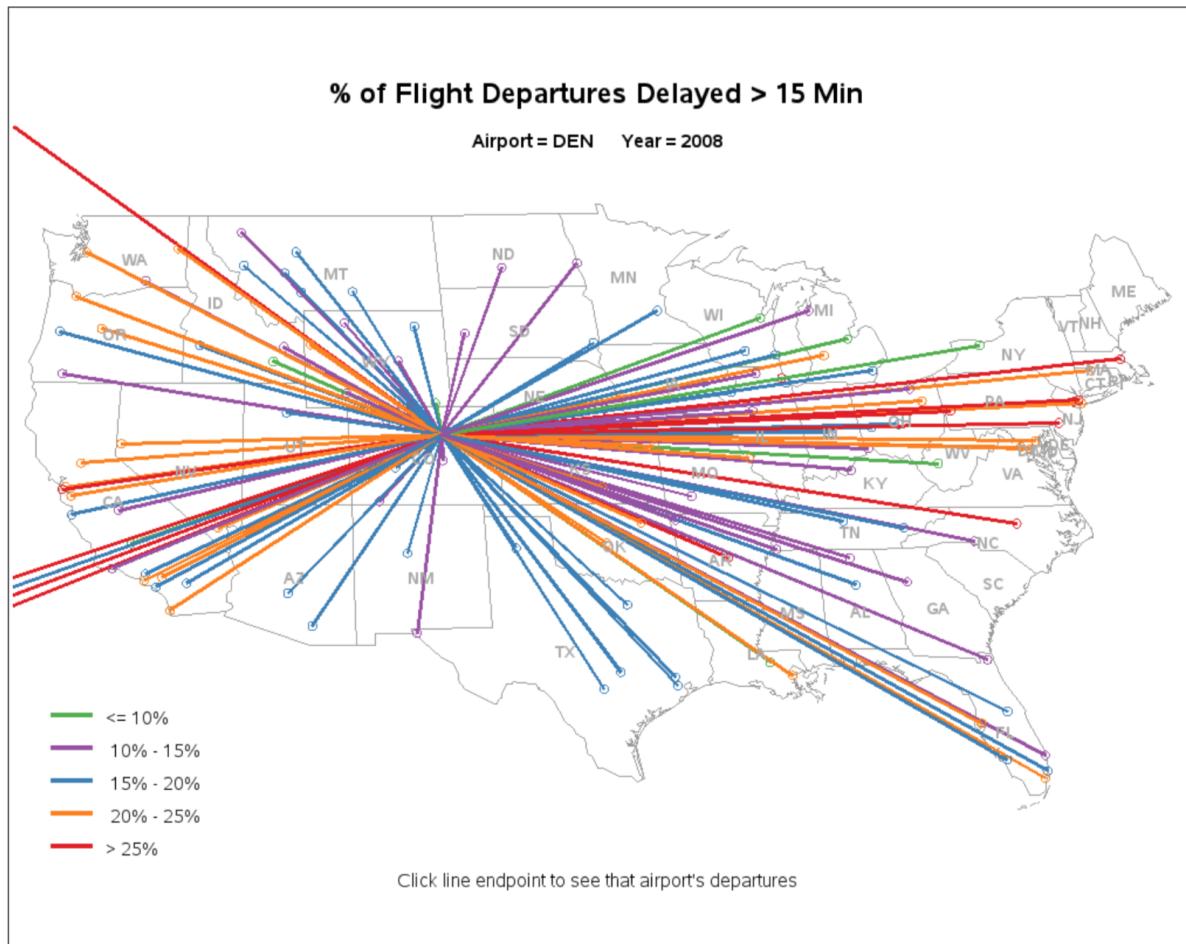
בתמונה קיימים שנים-עשר גרפים אשר מצינים את ממוצע האיחורים בכל חודש. רק באחד מן הגראפים מצויים הנתונים האמיתיים, בעוד בשאר הגראפים קיימם ערבול של הנתונים האמיתיים אשרណעד לדמות מצב אקרה בו החודשים בשנה אכן בלתוי תלויים. לאחר בחינה ניתן לחתות כי גורף מס' 10 הינו הגרף האמיתי. ניתן להראות את הגרף האמיתי על ידי צירטו שבולטת באתודות ועונתיות המראה קשר בין חלק מהחודשים הצמודים אשר לא ניתן לראות בסימולציה "אקרה".

אם לומדים מכך שכאשר אנו מבחינים בתוצאות בעלת תבנית מסוימת علينا לשאול את עצמנו, האם היא אקרה? בשל העובדה שגם יכולמים לזהות את הגרף הנכון מבין שאר הגראפים בעלי ההטפולוגיות האקראיות, אנו מעריכים שההטפואה אינה אקרה ומקורה בתופעה מסוימת. במקרה ההוסי הזה אנו לומדים שהשערת האפס השמרנית אינה נכונה, וכי קיימים סרדיים עונתיים. יתרה מכך, אנו לומדים על הכוח הרוב של עיצוב גרפי והיכולת שלו להזות מבחן סטטיסטי. מעבר ליכולת של הגרף לסתור סיפור ולהציג מידע כדי שראינו עד עתה בתרגום; אנו למדים על היכולת שלו לצרוב בגרף את התודעה שלנו והדעת הקדומות שאנו נשאים עמו על הנתונים. כך, שימושו במבחן זה עונה על השאלה "האם דעתנו על הנתונים היא ללא משוא-פנום?".

CODE

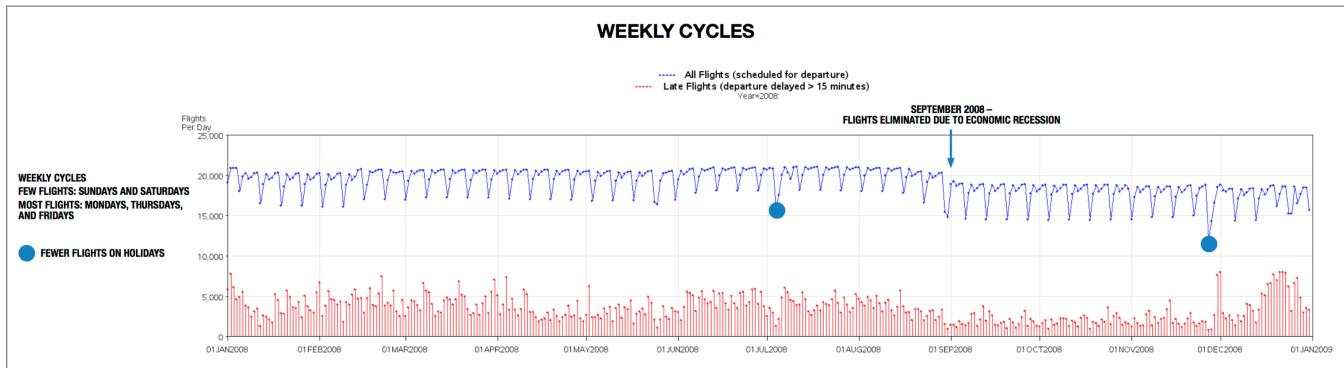
```
####<ins> שאלה ראשונה - ניתוח נתונים קיימים</ins>
```

```
library(magick)
frink <- image_read("https://i.imgur.com/lseR5HH.png")
image_trim(frink)
```

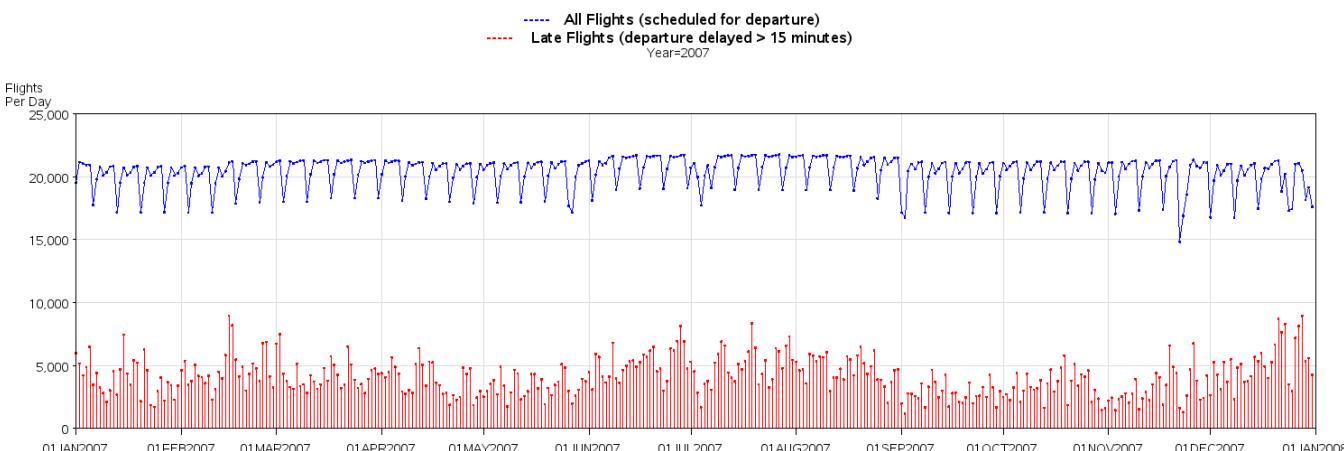


```
library(magick)
IM <- image_read("https://image.ibb.co/hmLnRx/Flight_Figure_2.png")
image_trim(IM)
```

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","fir
ebrick1")) %>%
  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 <br> LaGuardia Airport, 2013',
    geo = geo, showlegend = TRUE#, height=350
  )
)

```

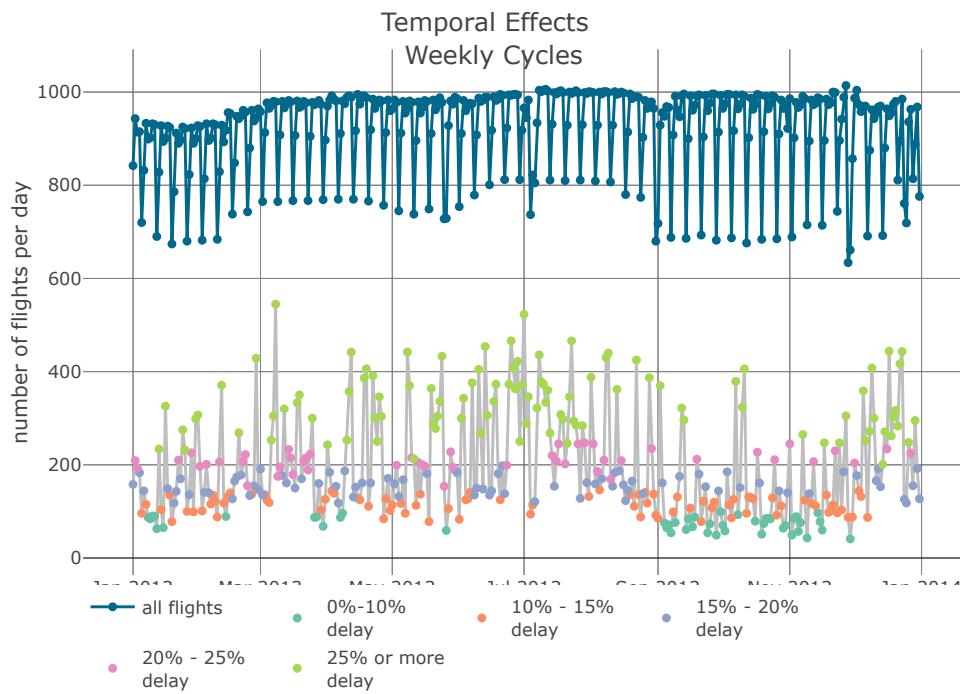


```

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$percent=A$delay/A$all

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

man_by_houer_plot <- ggplot(man_by_houer,aes(x=hour,y=percent,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 , angle=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")

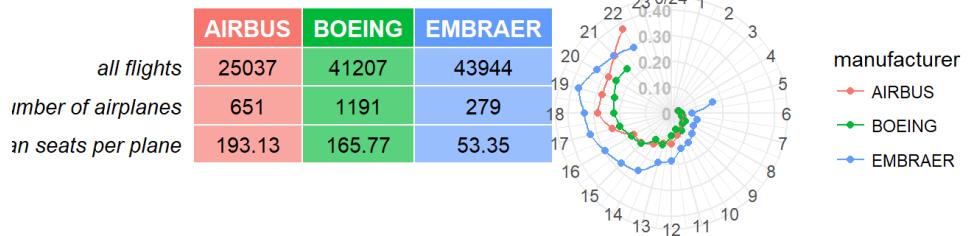
tab <- table_cell_bg(tab, row = 2, column = 4, linewidth = 1,
                     fill=rgb(red = 97/256,green = 156/256,blue = 255/256,alpha = 0.65), color = "white")
tab <- table_cell_bg(tab, row = 3, column = 4, linewidth = 1,
                     fill=rgb(red = 97/256,green = 156/256,blue = 255/256,alpha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

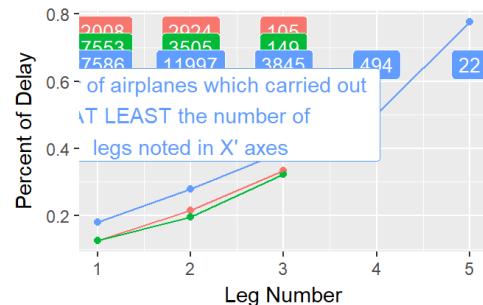
p1 <- ggarrange(tab,man_by_houer_plot)
p2 <- ggarrange(by_lag,man_by_air_time)
ggarrange(p1,p2,ncol=1,nrow=2,labels = "A View by Manufacturer, Airport EWR, Year 2013")

```

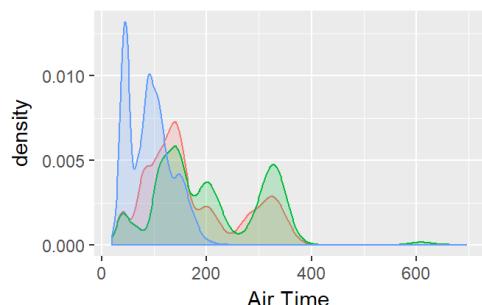
A View by Manufacturer, Airport and Year 2013 Throughout the Day



Percent of Flight Departures Delayed VS Leg Number



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 Precent")+
  xlab("Month")+
  facet_wrap(~sample_num)+
```

theme_minimal(base_size = 11, base_family = "")+

```

# scale_fill_brewer(palette="Spectral")
```

