

Exploratory Data Analysis

Flights and Airports

(Análisis de Vuelos y Aeropuertos)

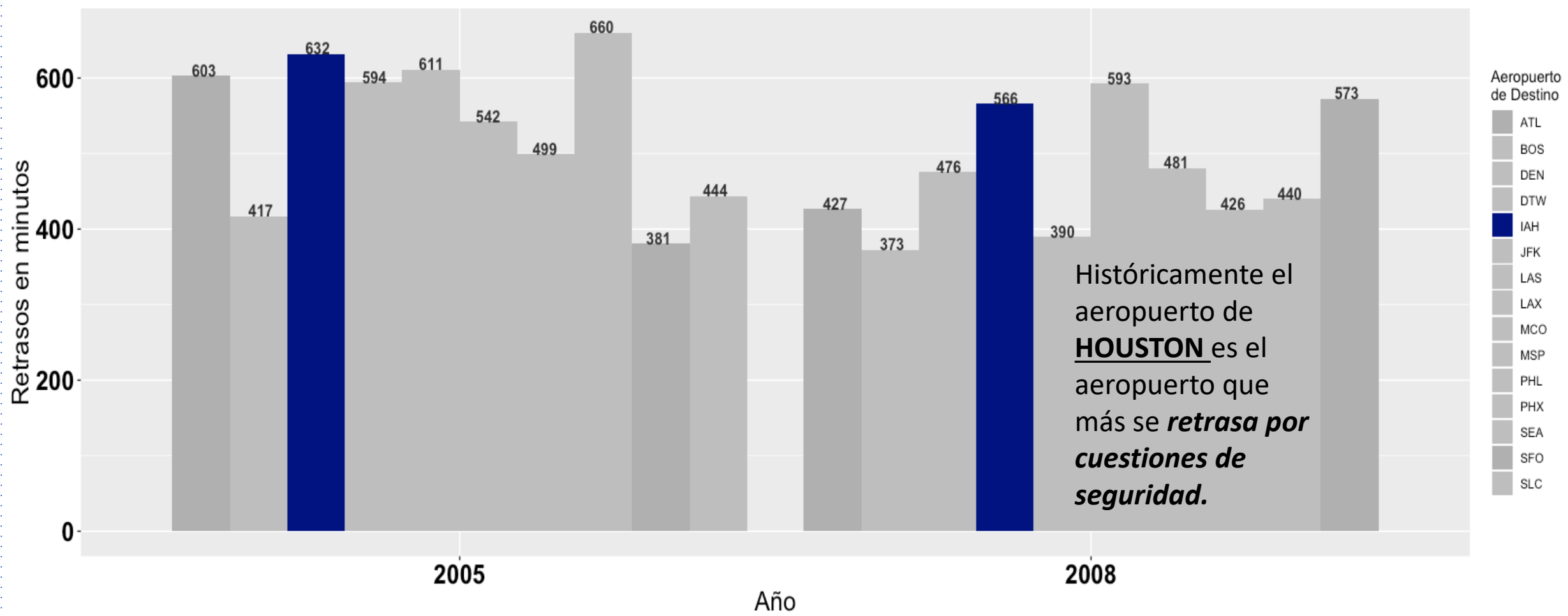
Andrés de la rosa

Objetivo del Análisis Exploratorio

Determinar las causas del retraso de seguridad en el aeropuerto de destino que ha experimentado históricamente el mayor retraso.

Top 10 de los Aeropuertos de destino más retrasados por Seguridad

Por año



No se registraba la variable retraso de seguridad en el 1990 ni en el 2000

El Retraso de Seguridad en el Aeropuerto de Destino Houston tiene una asociación con el aeropuerto de origen?

Grafico de Caja por Aeropuerto de Origen, en el Año 2005

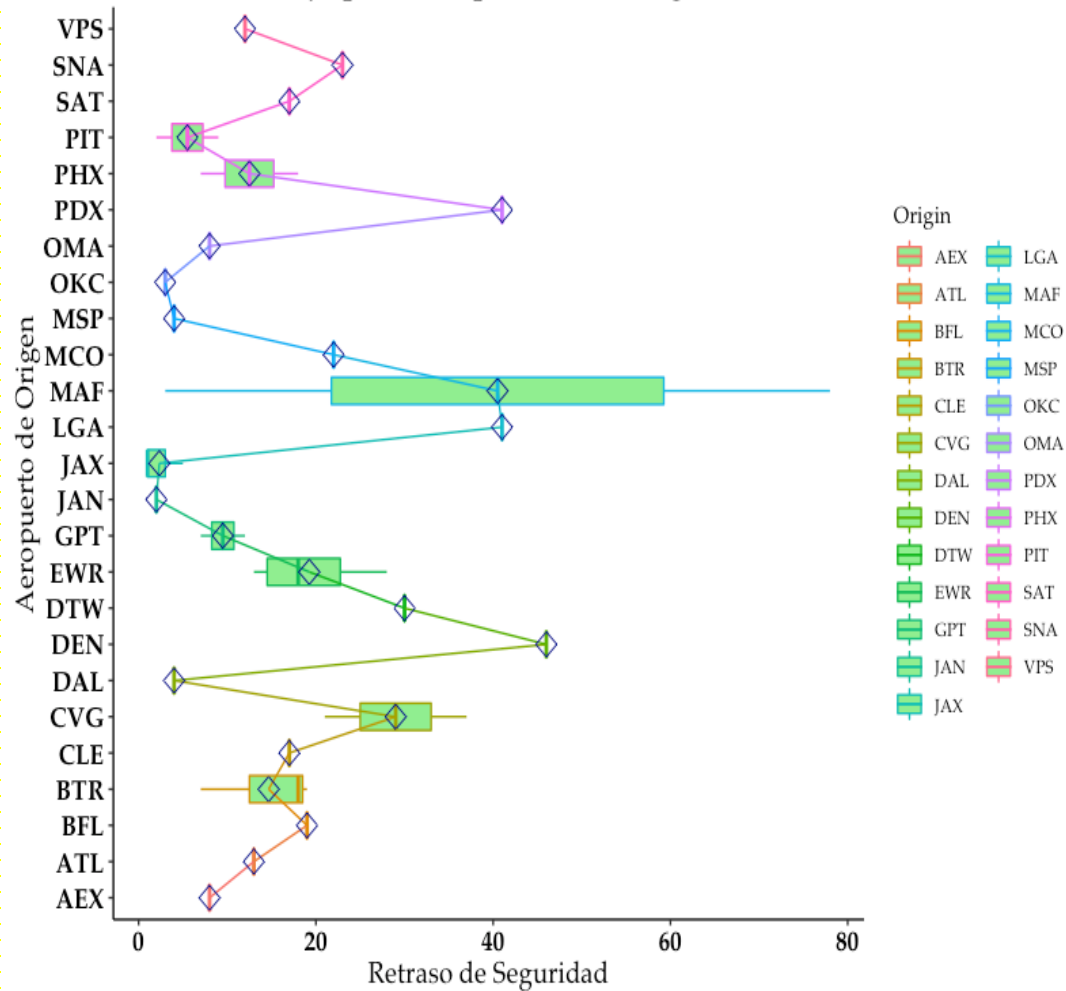
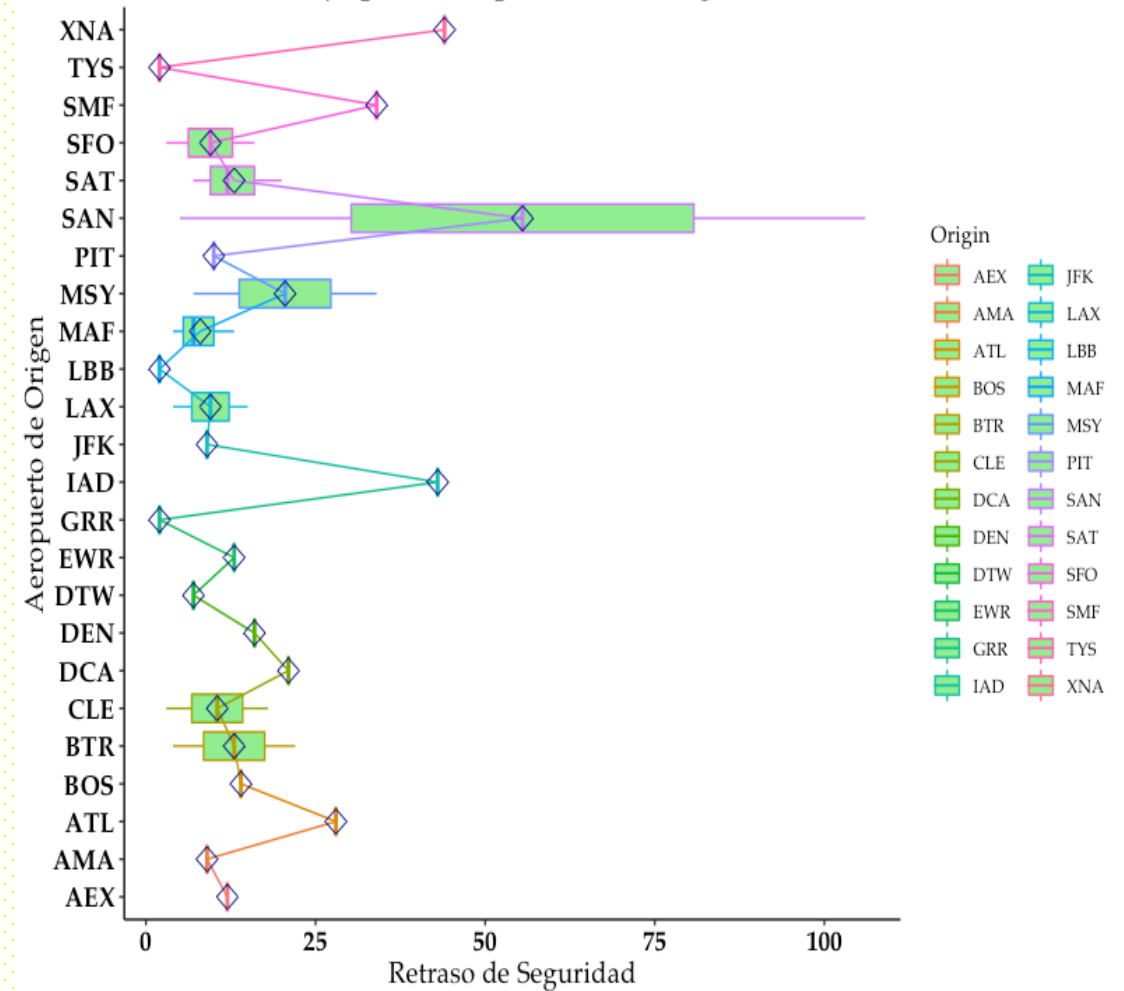


Grafico de Caja por Aeropuerto de Origen, en el Año 2008



Para los dos años, aunque se ven gráficamente las diferencias en las medias, las distribuciones de cada nivel dentro del factor aeropuerto de origen no son homogéneas (Varianza no homogénea) y no tienen el mismo n. Por otra parte, los aeropuertos de origen AEX, ATL, CLE, DEN, DTW, EWR, MAF, SAT, BTR Y PIT fueron recurrentes en los retrasos de seguridad para el 2005 y el 2008 y no existe un patrón definido con los otros aeropuertos de origen.

El Retraso de Seguridad en el Aeropuerto de Destino de Houston tiene una asociación con Aeropuerto de Origen?

Para estimar el efecto que tiene cada aeropuerto de origen en los retrasos de seguridad se modeló un One-way Anova para cada año. Dado que cada aeropuerto tiene apariciones distintas en esta data este es un modelo no balanceado. Entendiendo las implicaciones que existen si se incumple la condición de varianza homogénea. (varianza más grande de los parámetros estimados)

$$Retraso\ de\ Seguridad_{2005} = u + \beta_i AeroOrigen_{2005} + \varepsilon$$

$H_{0aeroOrigen}: \mu_{VPS} = \mu_{AEX} = \mu_{ATL} = \mu_{...}$ (El factor aeropuerto de origen no tiene un efecto en el retraso de seguridad del aeropuerto de Houston en el año 2005)

$H_{1aeroOrigen}: \mu_{VPS} \neq \mu_{AEX} \neq \mu_{ATL}$ (El factor Aeropuerto de origen tiene un efecto en el retraso de seguridad del aeropuerto de Houston en el año 2005, al menos uno de los niveles difiere)

Con un valor P de 0.376 no es posible rechazar la hipótesis nula. Por lo que el Aeropuerto de Origen no tiene un efecto en el retraso de seguridad ocurrido en el 2005 en el aeropuerto de Houston.

$Retraso\ de\ Seguridad_{2005} = u + \beta_i AeroOrigen_{2005} + \varepsilon$		
Suposición	Valor P	Conclusión
Normalidad (Shapiro- Wilk)	>0.000	No
Varianza Homogénea (Levene's)	>0.000	No
M>2 (número de replicas)	X	Modelo no balanceado
Resultado modelo alternativo (Kruskal-Wallis)	0.2612	Similar

$$Retraso\ de\ Seguridad_{2008} = u + \beta_i AeroOrigen_{2008} + \varepsilon$$

$H_{0aeroOrigen}: \mu_{VPS} = \mu_{AEX} = \mu_{ATL} = \mu_{...}$ (El factor aeropuerto de origen no tiene un efecto en el retraso de seguridad del aeropuerto de Houston en el año 2008)

$H_{1aeroOrigen}: \mu_{VPS} \neq \mu_{AEX} \neq \mu_{ATL}$ (El factor Aeropuerto de origen tiene un efecto en el retraso de seguridad del aeropuerto de Houston en el año 2008, al menos uno de los niveles difiere)

Con un valor P de 0.937 no es posible rechazar la hipótesis nula. Por lo que el Aeropuerto de Origen no tiene un efecto en el retraso de seguridad ocurrido en el 2008 en el aeropuerto de Houston.

$Retraso\ de\ Seguridad_{2008} = u + \beta_i AeroOrigen_{2008} + \varepsilon$		
Suposición	Valor P	Conclusión
Normalidad (Shapiro- Wilk)	>0.000	No
Varianza Homogénea (Levene's)	>0.000	No
M>2 (número de replicas)	X	Modelo no balanceado
Resultado modelo alternativo (Kruskal-Wallis)	0.7127	Similar

El Retraso de Seguridad en el Aeropuerto de Destino Houston tiene una asociación con la Aerolínea?

Grafico de Caja por Aerolinea, en el Año 2005

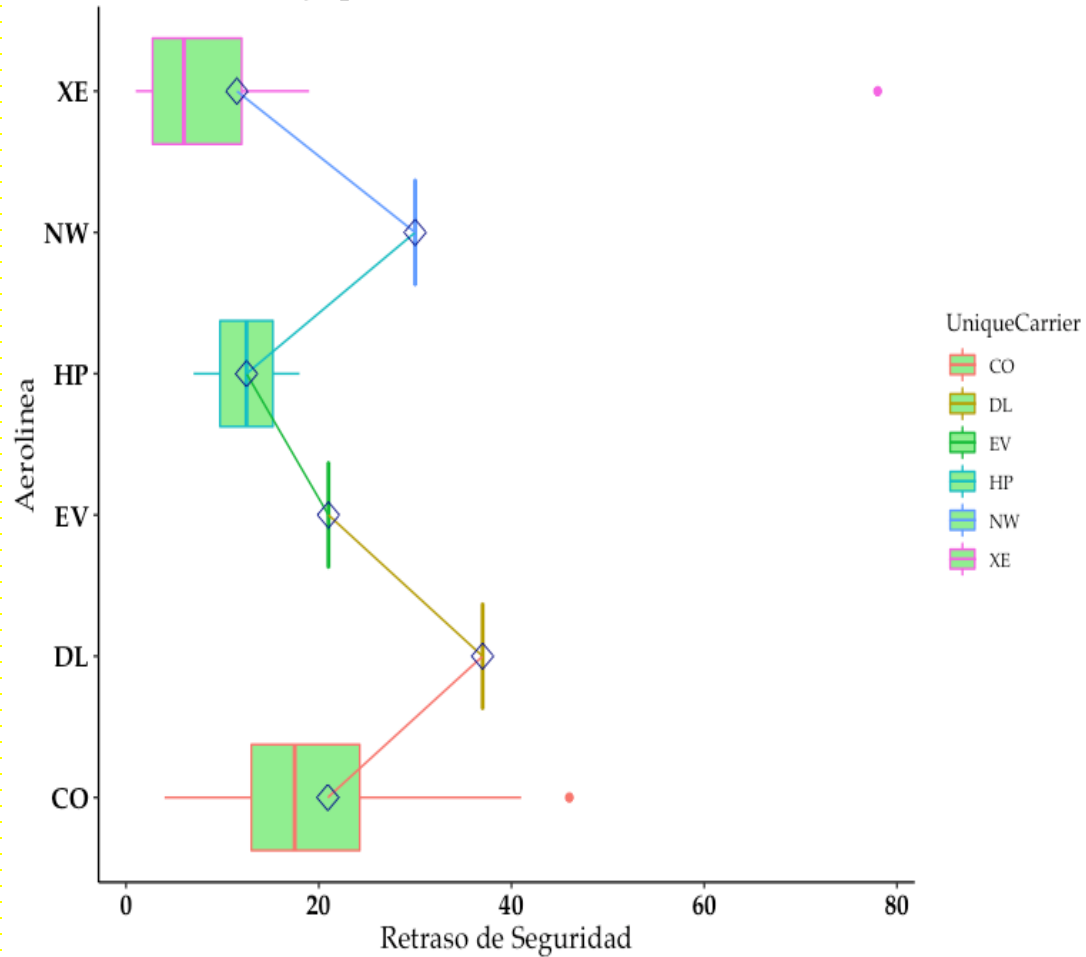
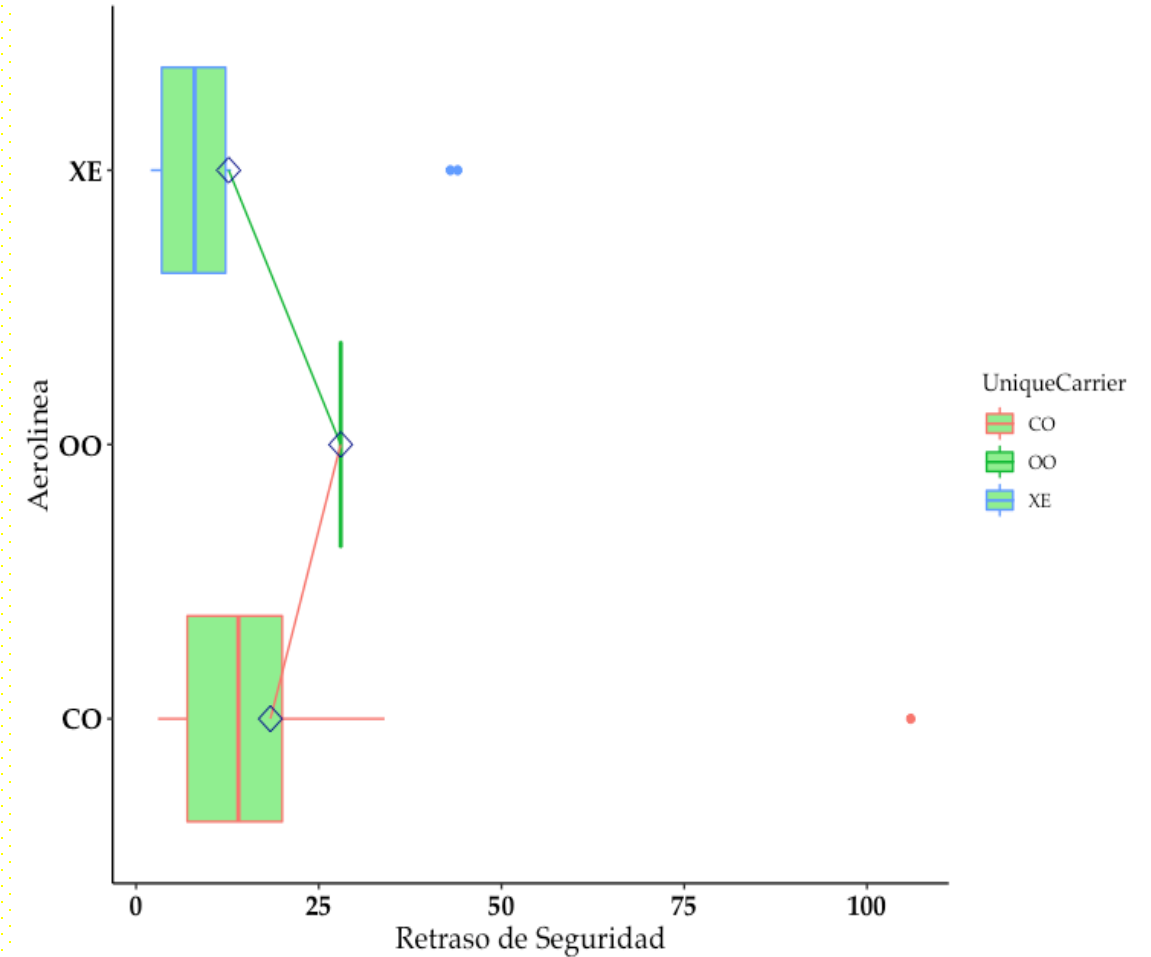


Grafico de Caja por Aerolinea, en el Año 2008



Para los dos años, aunque se ven gráficamente las diferencias en las medias, las distribuciones de cada nivel dentro del factor aerolínea no son homogéneas (Varianza no homogénea) y no tienen el mismo n. Por otra parte, las aerolíneas CO y XE fueron recurrentes en los retrasos de seguridad para el 2005 y el 2008.

El Retraso de Seguridad en el Aeropuerto de Destino de Houston tiene una asociación con la Aerolínea?

Para estimar el efecto que tiene cada aerolínea en los retrasos de seguridad se modeló un One-way Anova para cada año. Dado que cada aerolínea tiene apariciones distintas en esta data este es un modelo no balanceado. Entendiendo las implicaciones que existen si se incumple la suposición de varianza constante.

$Retraso\ de\ Seguridad_{2005} = u + \beta_i Aerolinea_{2005} + \varepsilon$

$H_{0Aerolinea}$: $\mu_{XE} = \mu_{OO} = \mu_{CO}$ (El factor aerolínea no tiene un efecto en el retraso de seguridad del aeropuerto de Houston en el año 2005)

$H_{1Aerolinea}$: $\mu_{XE} \neq \mu_{OO} \neq \mu_{CO}$ (El factor aerolínea tiene un efecto en el retraso de seguridad del aeropuerto de Houston en el año 2005, al menos uno de los niveles difiere)

Con un valor P de 0.376 no es posible rechazar la hipótesis nula. Por lo que la Aerolínea no tiene un efecto en el retraso de seguridad ocurrido en el 2005 en el aeropuerto de Houston.

$Retraso\ de\ Seguridad_{2005} = u + \beta_i Aerolinea_{2005} + \varepsilon$		
Suposición	Valor P	Conclusión
Normalidad (Shapiro- Wilk)	>0.000	No
Varianza Homogénea (Levene's)	0.9324	Si
M>2 (número de replicas)	X	Modelo no balanceado
Resultado modelo alternativo (Kruskal-Wallis)	0.02325	Tiene un Efecto al 5%

$Retraso\ de\ Seguridad_{2008} = u + \beta_i Aerolinea_{2008} + \varepsilon$

$H_{0Aerolinea}$: $\mu_{XE} = \mu_{OO} = \mu_{CO}$ (El factor aerolínea no tiene un efecto en el retraso de seguridad del aeropuerto de Houston en el año 2008)

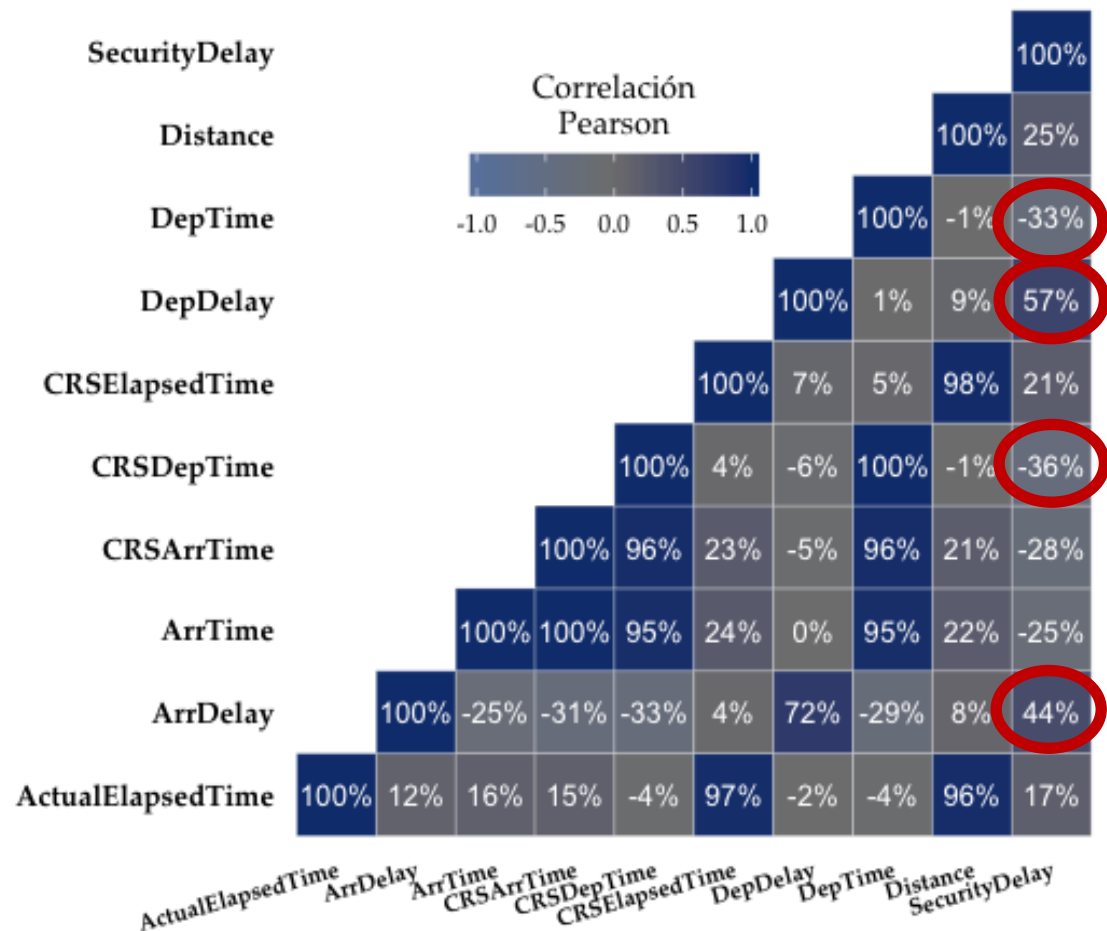
$H_{1Aerolinea}$: $\mu_{XE} \neq \mu_{OO} \neq \mu_{CO}$ (El factor aerolínea tiene un efecto en el retraso de seguridad del aeropuerto de Houston en el año 2008, al menos uno de los niveles difiere)

Con un valor P de 0.617 no es posible rechazar la hipótesis nula. Por lo que la Aerolínea no tiene un efecto en el retraso de seguridad ocurrido en el 2008 en el aeropuerto de Houston.

$Retraso\ de\ Seguridad_{2008} = u + \beta_i Aerolinea_{2008} + \varepsilon$		
Suposición	Valor P	Conclusión
Normalidad (Shapiro- Wilk)	>0.000	No
Varianza Homogénea (Levene's)	0.8024	Si
M>2 (número de replicas)	X	Modelo no balanceado
Resultado modelo alternativo (Kruskal-Wallis)	0.1575	Similar al 10%

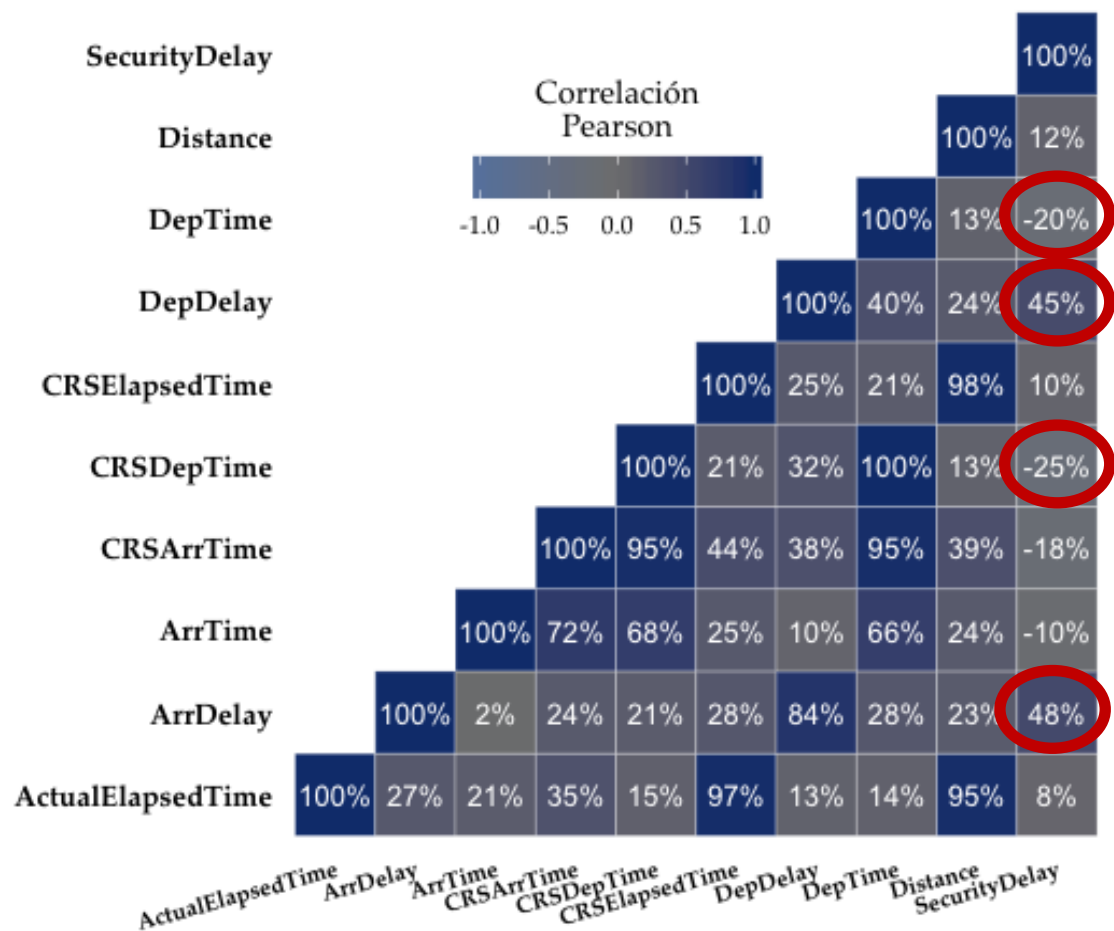
El Retraso de Seguridad en el Aeropuerto de Destino de Houston se mueve en conjunto con otra variable cuantitativa de este dataset ?

Mapa de Calor y Correlaciones
Buscando Relaciones con el Retraso de Seguridad
En el Aeropuerto de Houston en el año 2005



Las Variables Weather Delay y Carrier Delay están vacías para esta muestra

Mapa de Calor y Correlaciones
Buscando Relaciones con el Retraso de Seguridad
En el Aeropuerto de Houston en el año 2008



Las Variables Weather Delay y Carrier Delay están vacías para esta muestra

No existe una variable continua que este fuertemente correlacionada con el retraso de seguridad del aeropuerto de Houston. Existen algunas señaladas en la matriz que requieren una análisis que aprenderemos más adelante en este curso.

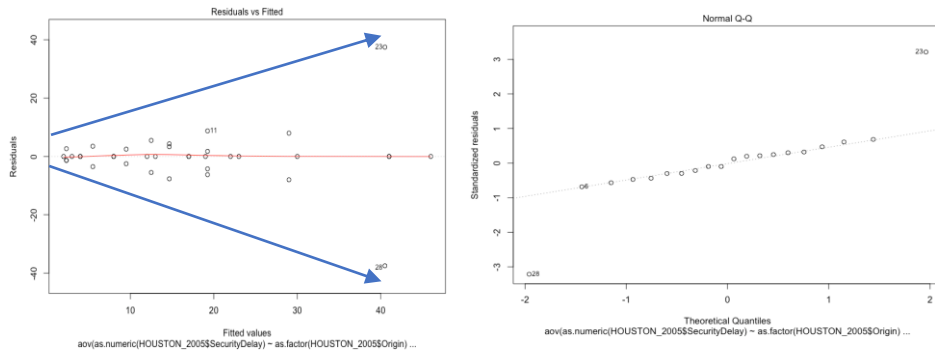
Conclusiones

- Basados en los modelos One- Way Anova contruidos, los retrasos en el Aeropuerto de Houston ocurridos en el año 2005 y el 2008, no guardan una asociación con el Aeropuerto de origen ni con la Aerolínea. Para controlar los sesgos se utilizó como modelo alternativo pruebas de Kruskal-Wallis y estas arrojan un resultado similar para todos los casos menos en el modelo de Aerolínea del 2005, a donde este modelo sugiere que la Aerolínea tiene un efecto en el retraso de seguridad al 5%.
- Basado en una matriz de correlación no existe una variable que este fuertemente correlacionada con el retraso de seguridad del aeropuerto de HOUSTON.

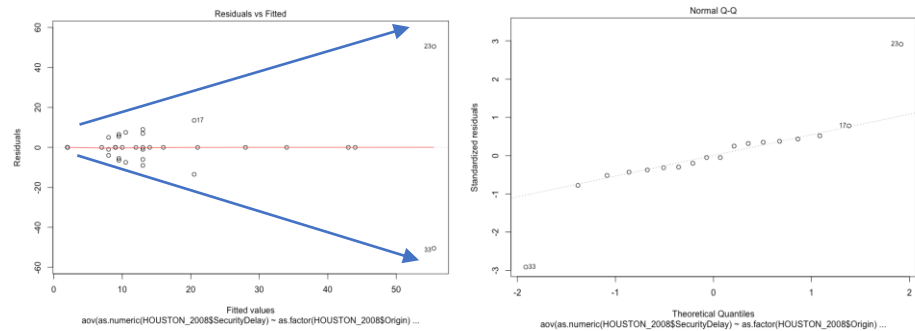
Anexo- Tests y gráficos de Diagnóstico de los Modelos

Modelos Aeropuerto de Origen

2005

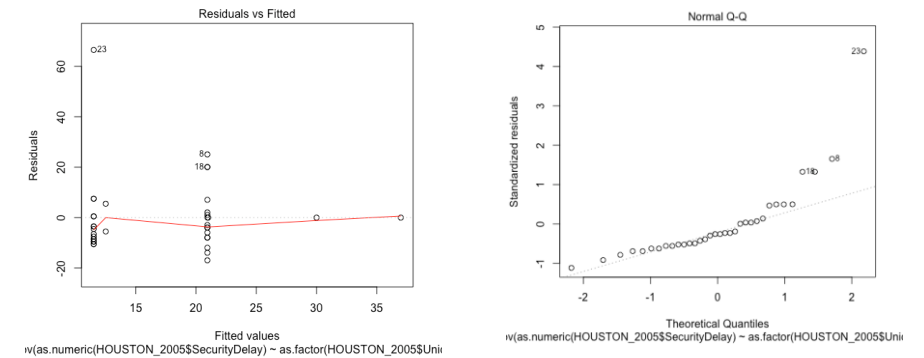


2008



Modelos Aerolínea

2005



2008

