

Hoy te convertís en (Junior) Data Scientist

Universidad de Buenos Aires
Buenos Aires, Argentina

- Cortés Lucas
- Lamela Emanuel
- Zimenspitz Ezequiel



1. Introducción

Motivación:

- Analizar datos reales de vuelos realizados en USA
- Identificar modelos que describan algunos comportamientos y evaluar su eficacia para realizar predicciones
- *Data fitting* a través de **cuadrados mínimos**

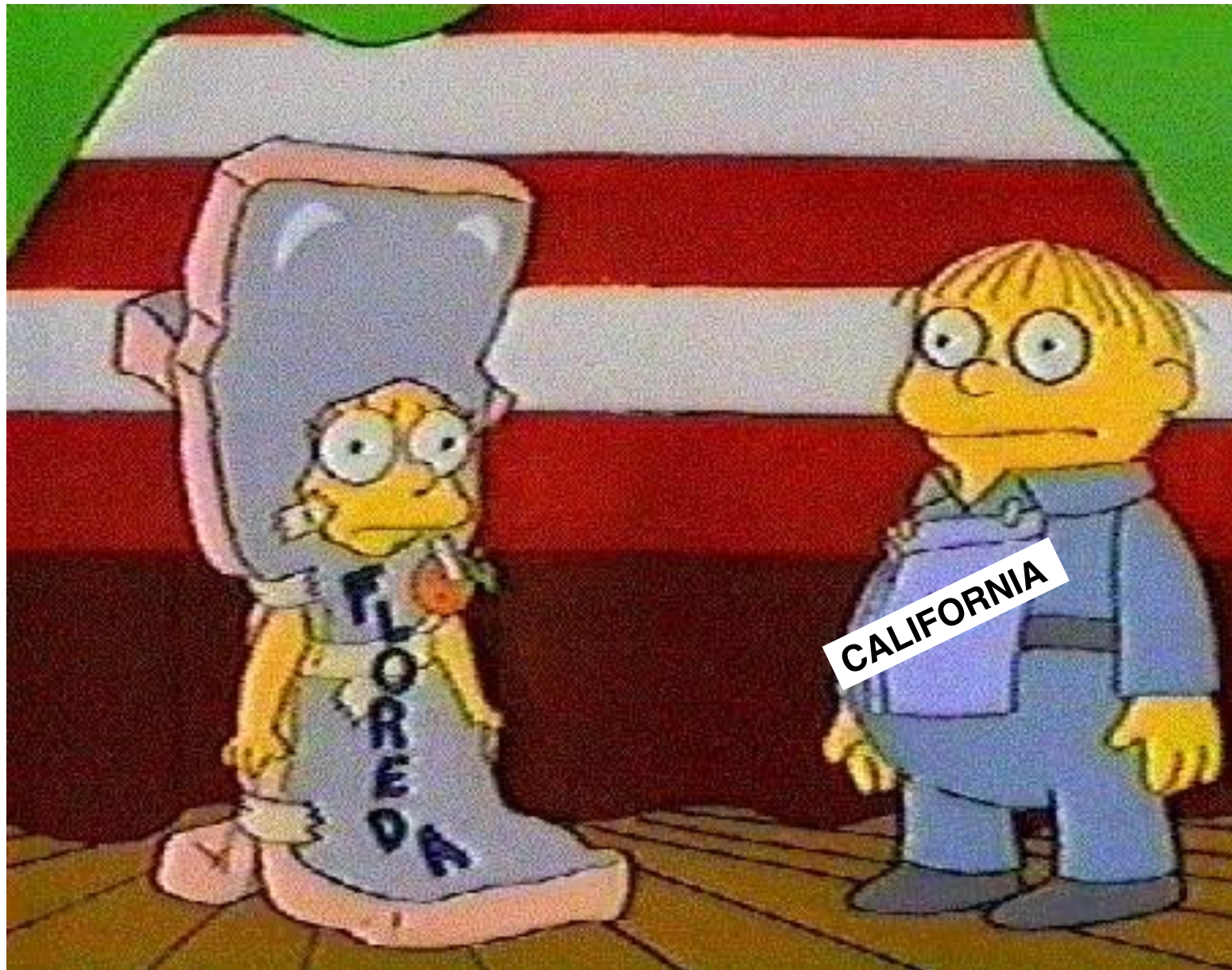
2. Primeros pasos

- 12Gb de datos entre los años '87 a '08
- Restricción del 2000 al 2008
- Planteo de ejes de estudio
 - Desvíos de vuelos en relación al **tiempo**
 - Desvíos de vuelos en relación a la **distancia**

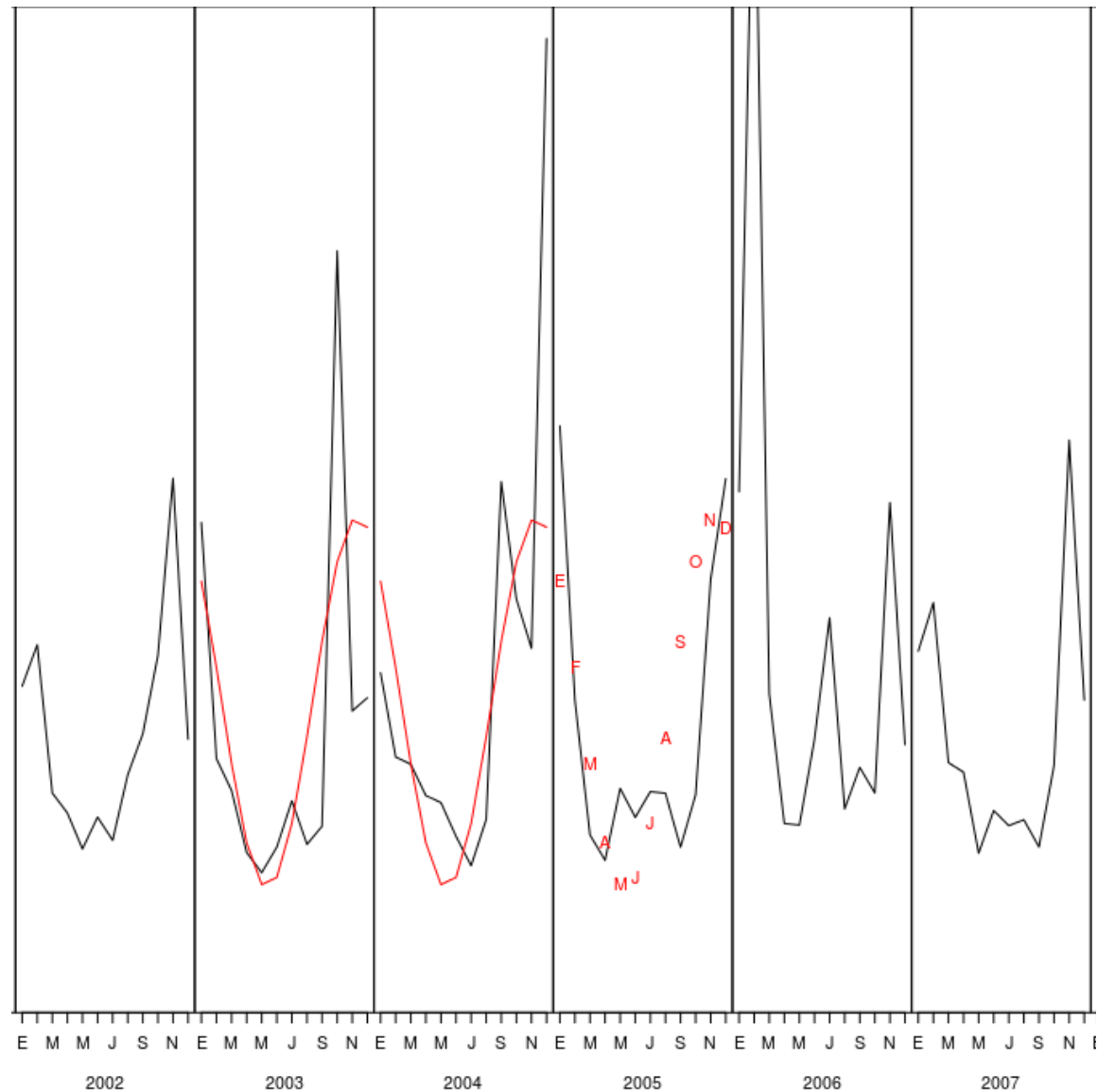
3. Experimentación

*Desvíos de vuelos en relación al **tiempo***

California vs Florida



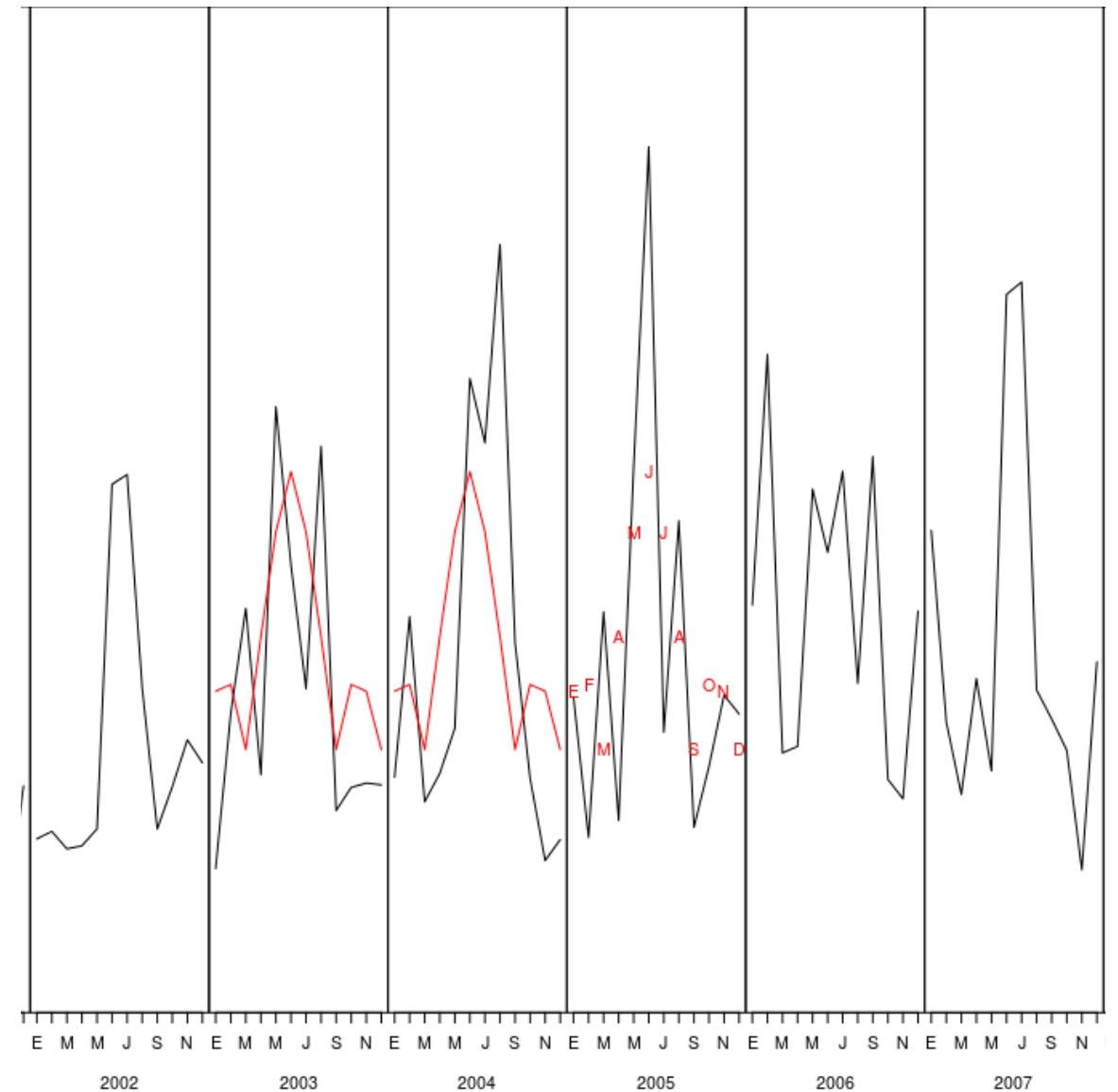
California



$$F_1 = \alpha_1 * \text{abs}(\sin(\frac{\pi}{12} * x) * \cos(\frac{\pi}{6} * x)^2) + \alpha_2$$

● $F_2 = \alpha_1 * \sin(\frac{\pi}{6} * x) + \alpha_2 * \cos(\frac{\pi}{6} * x) + \alpha_3$

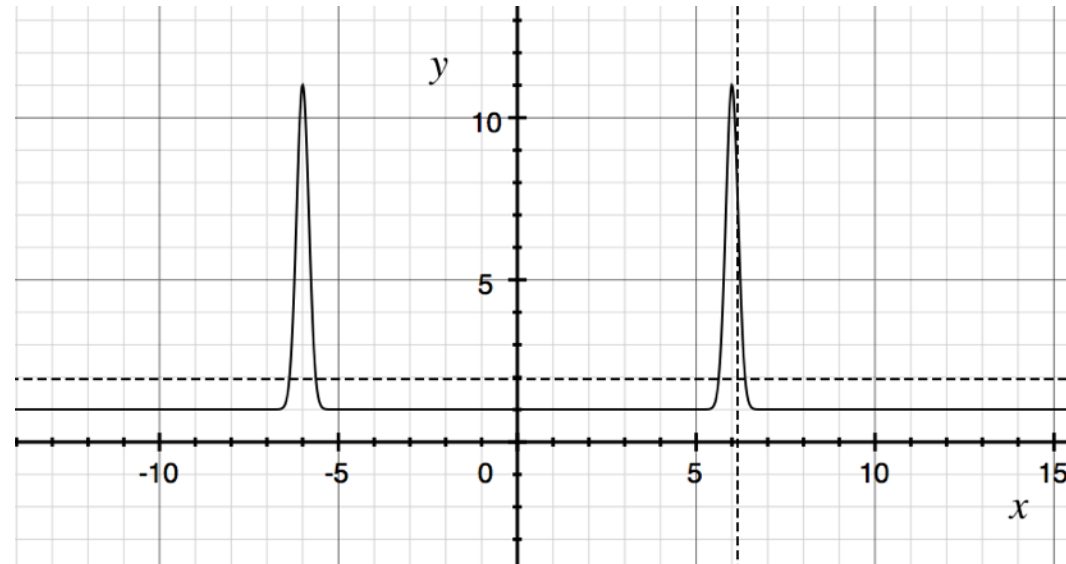
Florida



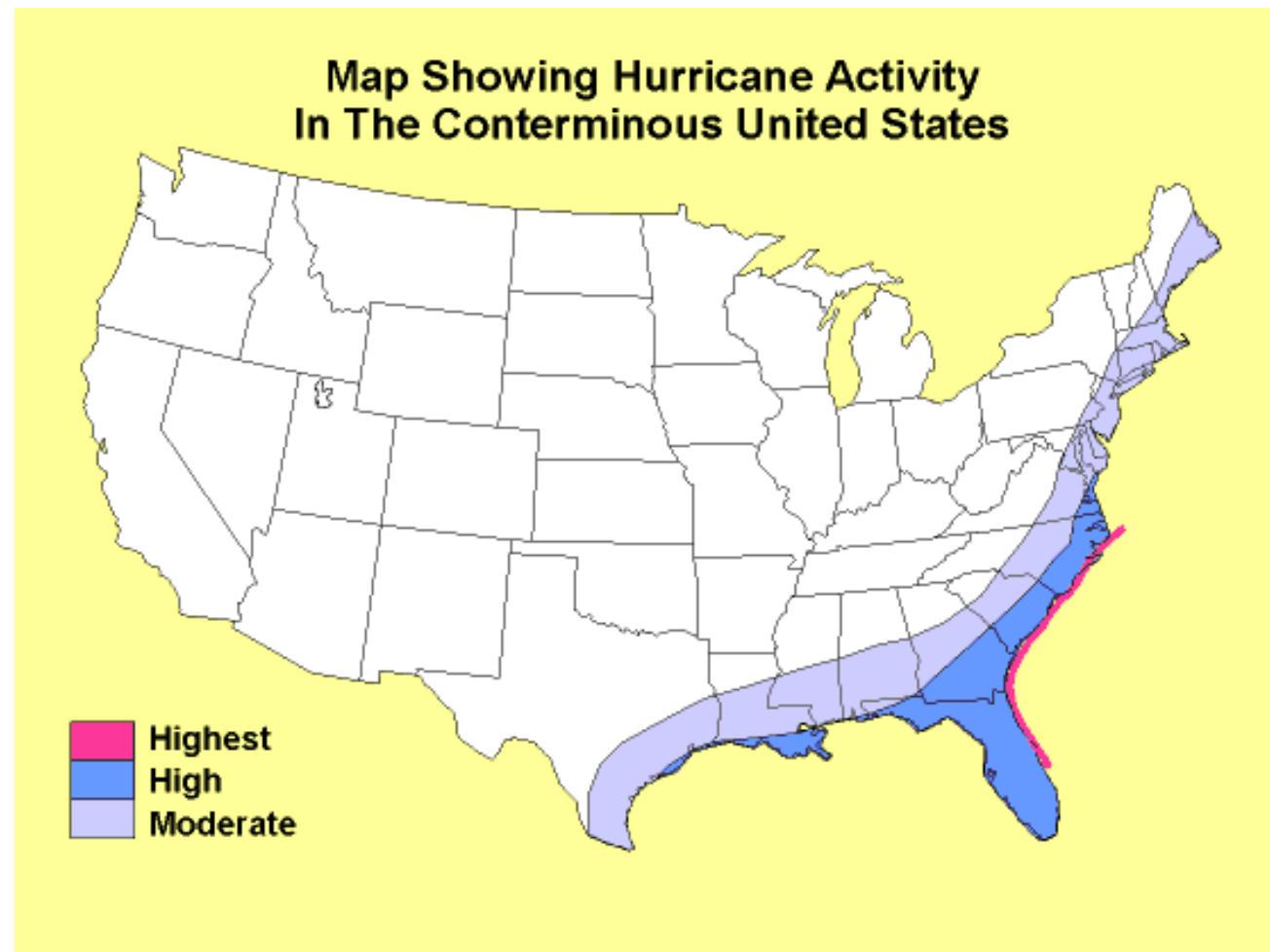
$$F_1 = \alpha_1 * \text{abs}(\sin(\frac{\pi}{12} * x) * \cos(\frac{\pi}{6} * x)^2) + \alpha_2$$

$$F_2 = \alpha_1 * \sin(\frac{\pi}{6} * x) + \alpha_2 * \cos(\frac{\pi}{6} * x) + \alpha_3$$

● $F_3 = \alpha_1 * \text{abs}(\sin(\frac{\pi}{12} * x) * \cos(\frac{\pi}{6} * x)^2) +$
 $+ \alpha_2 * \cos(\frac{\pi}{12} * x - \frac{\pi}{2})^{500} + \alpha_3$

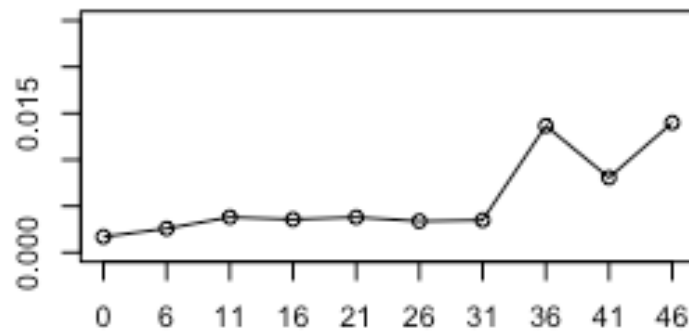


$$F = 5 * \cos(x)^{500} + 1$$

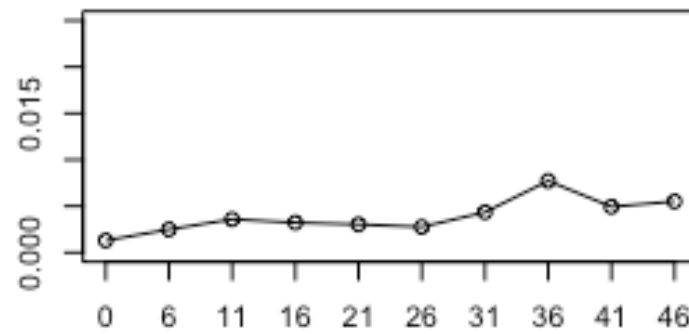


3. Experimentación

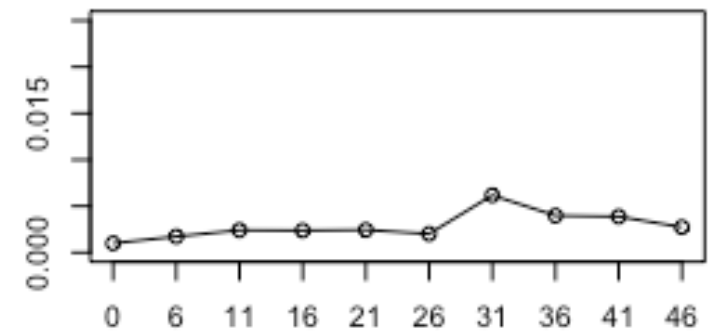
*Desvíos de vuelos en relación a la **distancia***



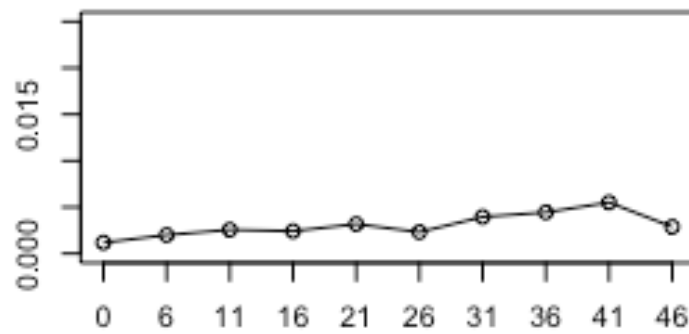
2000



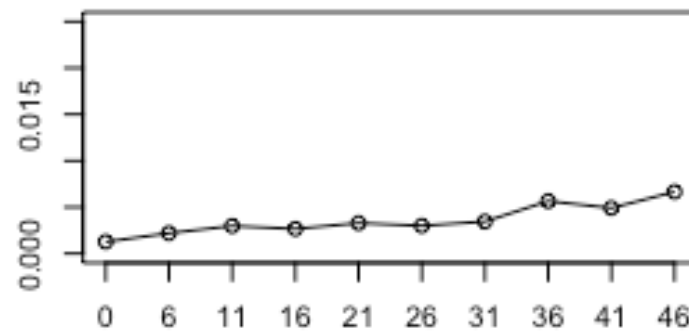
2001



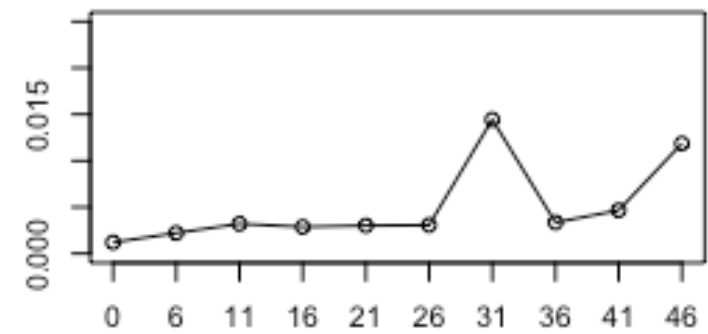
2002



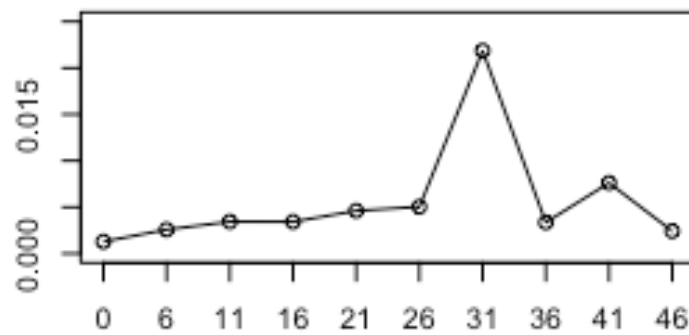
2003



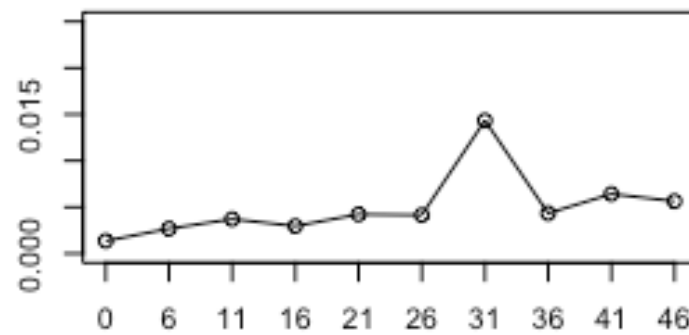
2004



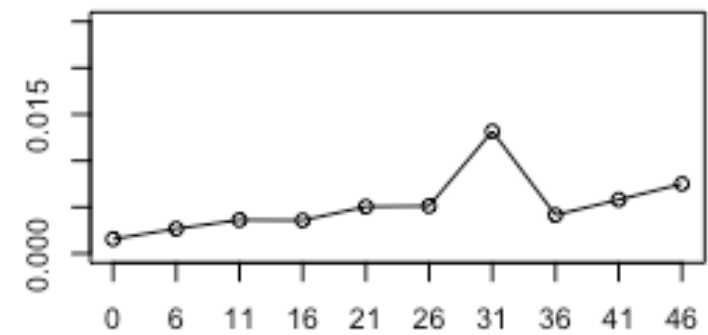
2005



2006



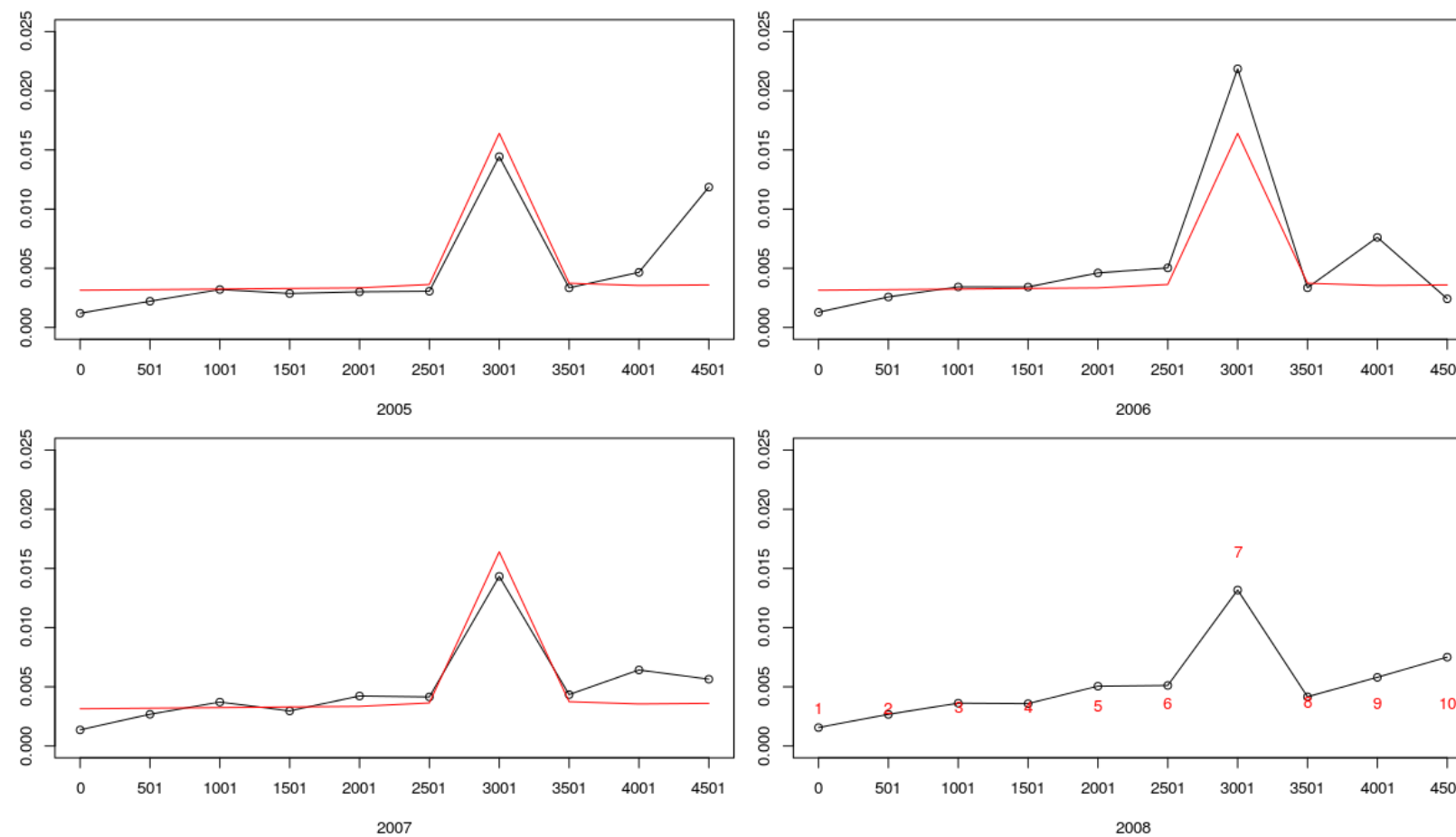
2007



2008

3. Experimentación

*Desvíos de vuelos en relación a la **distancia***

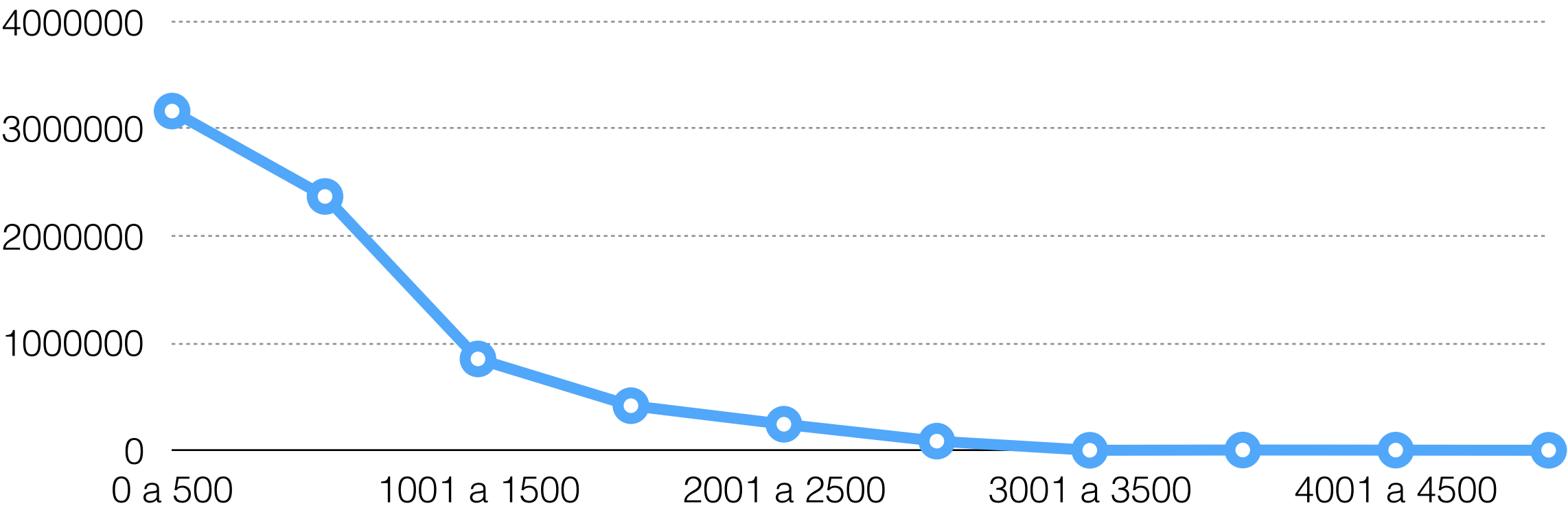


$$F_1 = \alpha_1 * x + \alpha_2$$

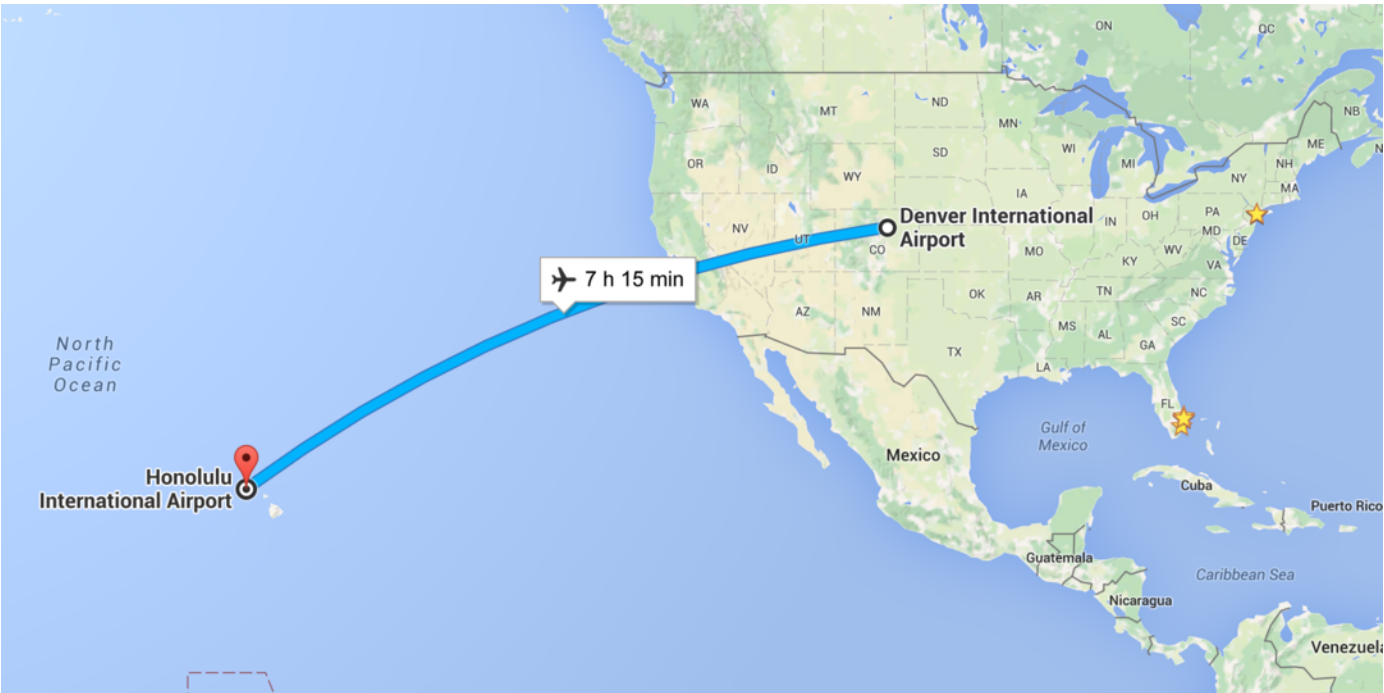
$$F_2 = \alpha_1 * x^2 + \alpha_2 * x + \alpha_3$$

$$\bullet F_3 = \alpha_1 * e^{-\frac{(i-7)^2}{(2*0.25)^2}} + \alpha_2 * x + \alpha_3$$

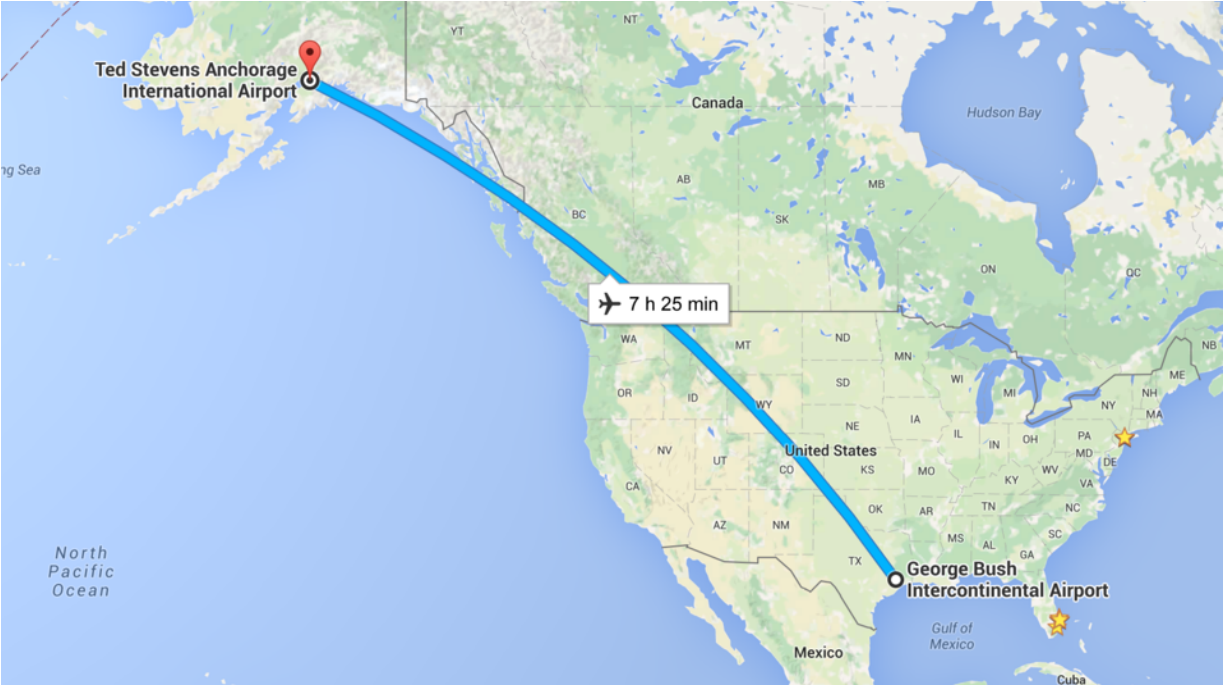
Flights qty per distance range



DEN <-> HNL



IAH <-> ANC



4. Conclusiones y trabajo futuro

- Cuadrados mínimos lineales efectivos pero depende fuertemente de la familia de funciones elegida.
- Hay una relación entre los desvíos y la época del año / distancia.



- ◉ Aplicar los modelos predictivos en otros Estados, por ejemplo NY, que presentan otros factores particulares -por ejemplo, climatológicos-
- ◉ En base al análisis hecho ver como se pueden reducir los desvíos.

Gracias!
Preguntas?

