jet

jetBlue

# R Makes the World Go ‘Round:
# **Data-Driven Decision Making at JetBlue**

Catherine Zhou
NY R Conference
April 21, 2017

# New York's Hometown Airline

- Headquartered in Long Island City
- Focus Cities: Boston, Ft Lauderdale, Long Beach, New York, Orlando, San Juan
- U.S., Caribbean, and Latin America
- Havana, Cuba in Nov 2016 - 100th city
- **Inaugural flight on Feb 11, 2000**

# FINANCIAL CONDITION OF THE AIRLINE INDUSTRY

I don't think JetBlue has a better chance of being profitable than 100 other predecessors with new airplanes, new employees, low fares, all touchy-feely… all of them are losers. Most of these guys are smoking ragweed.

Gordon Bethune, CEO Continental Airlines, June 2002

**Introduction to Air Transport Economics: From Theory to Applications**
Bijan Vasigh, Ken Fleming

# US MAJOR AIRLINES SINCE 1950

***Source:*** *Aviation Week Intelligence Network's Fleet Database, airline websites*

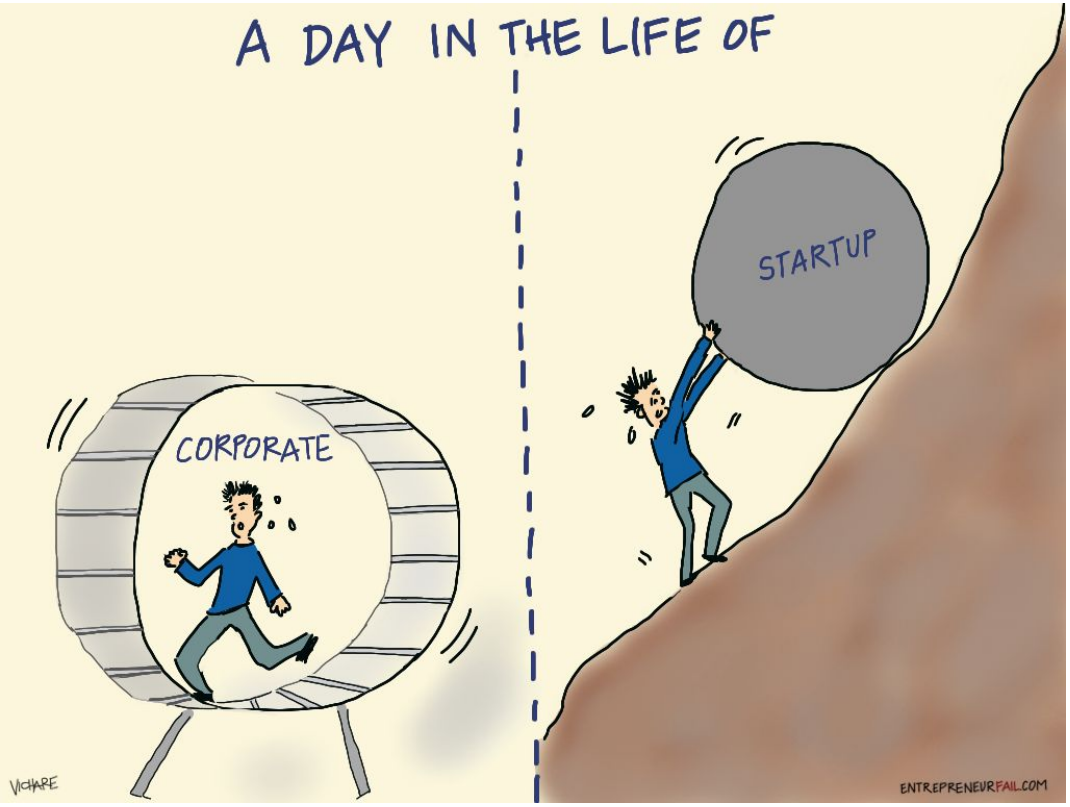
## 2015 SCHEDULED DOMESTIC AND INTERNATIONAL ENPLANEMENTS (U.S. CARRIERS)

| Rank | Carrier | Enplaned Passengers |
|---|---|---|
| 1 | American/U.S. Airways | 146.55 |
| 2 | Southwest | 144.575 |
| 3 | Delta | 138.591 |
| 4 | United | 95.327 |
| **5** | **JetBlue** | **35.074** |
| 6 | SkyWest | 30.078 |
| 7 | ExpressJet | 26.003 |
| 8 | Alaska | 22.828 |
| 9 | Spirit | 17.629 |
| 10 | Republic | 13.908 |

Passenger numbers in millions (000,000)

Source: Bureau of Transportation Statistics, T-100 Market

# Airline with growing pains and desire to innovate



# Data Challenges

- Legacy software and lack of data integration
- No core data team (can be pro and con)
- Speaking across workgroups
- Data architecture is dated
- Scaling analysis for a growing organization

# Airline Data is Dream Data

On a structural level:

- Much of our data is reported publicly
- Standardized -- fairly well structured
- Lots of data. Lots of it.

On a personal level:

- Analyzing movement patterns
- Deciding where and how people fly
- A transportation nerd's dream!

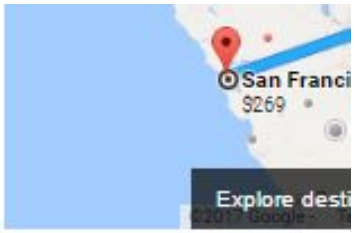
# “When is the best time to buy a plane ticket?”

JFK New York City +

SFO San Francisco +

Thu, April 27

Add return date




One way flight

Thu, Apr 27

jetBlue

**7:00 AM – 10:44 AM**
New York City (JFK) – San Francisco (SFO)
JetBlue 15 · Economy Class · Airbus A320
Above average legroom (33")

6h 44m
Wi-Fi
In-seat & USB power
On-demand video

Book, track or share this flight
Total incl taxes & fees for 1 adult · Additional bag fees may apply.

Ads

Book with **JetBlue** $269

Track price
Save this trip to track price changes and receive price alerts and travel tips by email. Learn more

Track price

10% fare increase likely in 8h · Show less

The currently bookable fare for this flight expires soon. The new fare will likely be more expensive. Historically, expected fare increases on similar flights have been accurate 94% of the time.

Matrix Airfare Search

## Fare rules & restrictions

JetBlue (B6) MH7JUEN5 NYC to SFO

### General notes

### Category 4: Flight restrictions

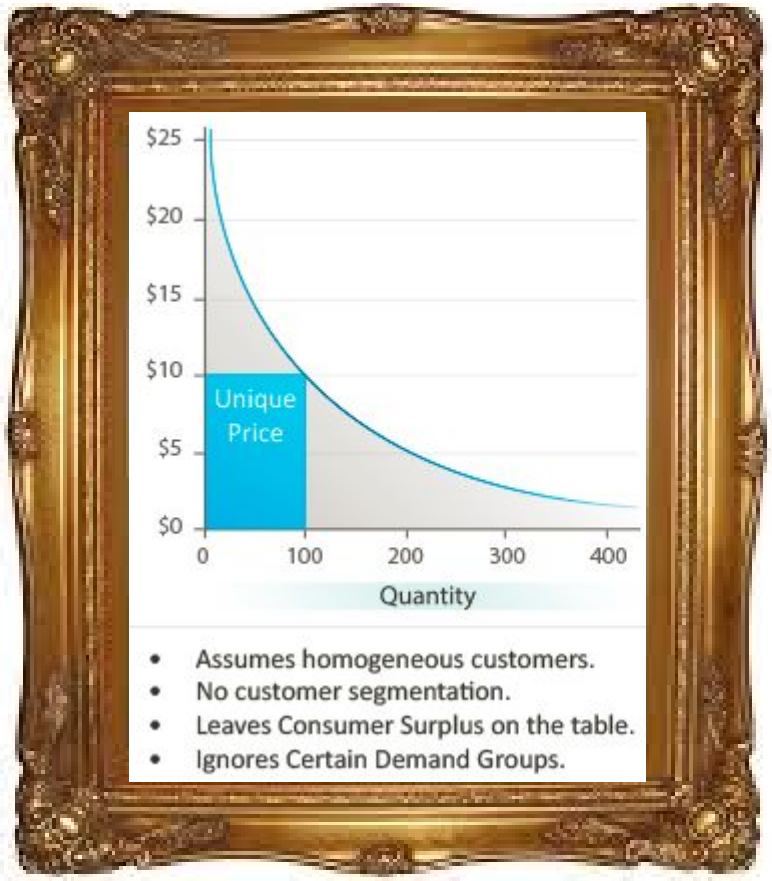
```
THE FARE COMPONENT MUST NOT BE ON
    ONE OR MORE OF THE FOLLOWING
      B6 FLIGHTS 5700 THROUGH 5799 OPERATED BY BB
      ANY B6 FLIGHT OPERATED BY 9K.
AND
THE FARE COMPONENT MUST BE ON
    ONE OR MORE OF THE FOLLOWING
      B6 FLIGHTS 0001 THROUGH 3999 OPERATED BY B6
      B6 FLIGHTS 8000 THROUGH 8499 OPERATED BY B6.
```

### Category 5: Advanced reservation/ticketing restrictions

```
RESERVATIONS FOR ALL SECTORS ARE REQUIRED AT LEAST 7 DAYS
BEFORE DEPARTURE.
```

# Brief History of Revenue Management

## Classic Airline Pricing before 1980s




- Assumes homogeneous customers.
- No customer segmentation.
- Leaves Consumer Surplus on the table.
- Ignores Certain Demand Groups.

## Airline RM 1980s-present

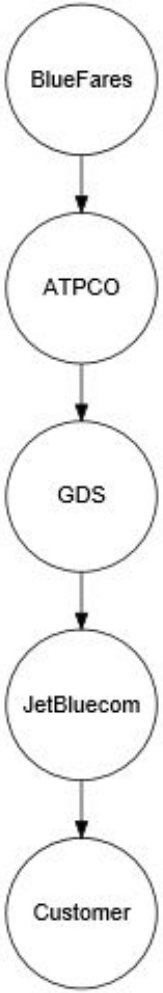
- Assumes heterogeneous customers.
- Multiple segmentation.
- Gains consumer´s surplus.
- More satisfied demand.

# Pricing and Inventory

- Pricing determines dollar amount for fares
- Pricing determines fare rules
  - Advanced purchase restrictions
  - First travel/last travel
  - Day of week

- Inventory allocates seats available in each fare class

# Lifecycle of a Fare



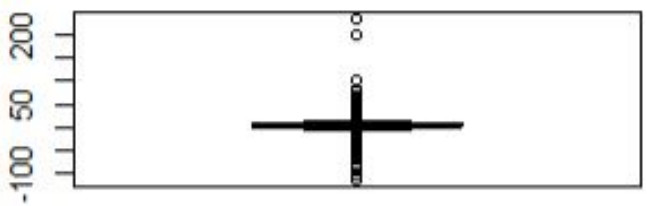

- Fare changes are sent to a clearinghouse for government filing system
  - Air Tariff Publishing Company (ATPCO)
- Fares made public through Global Distribution System
  - Created during era of paper tickets & travel agents
- GDS sends fares to airline websites and OTAs
  - Platform distributes fare and inventory data to airline websites and online travel agencies
- Pricing team analyzes competitive fare changes throughout the day

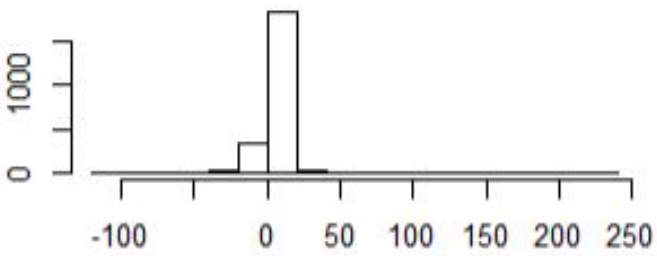
REUTERS U.S. airlines raise fares $3 in first sector-wide hike since June

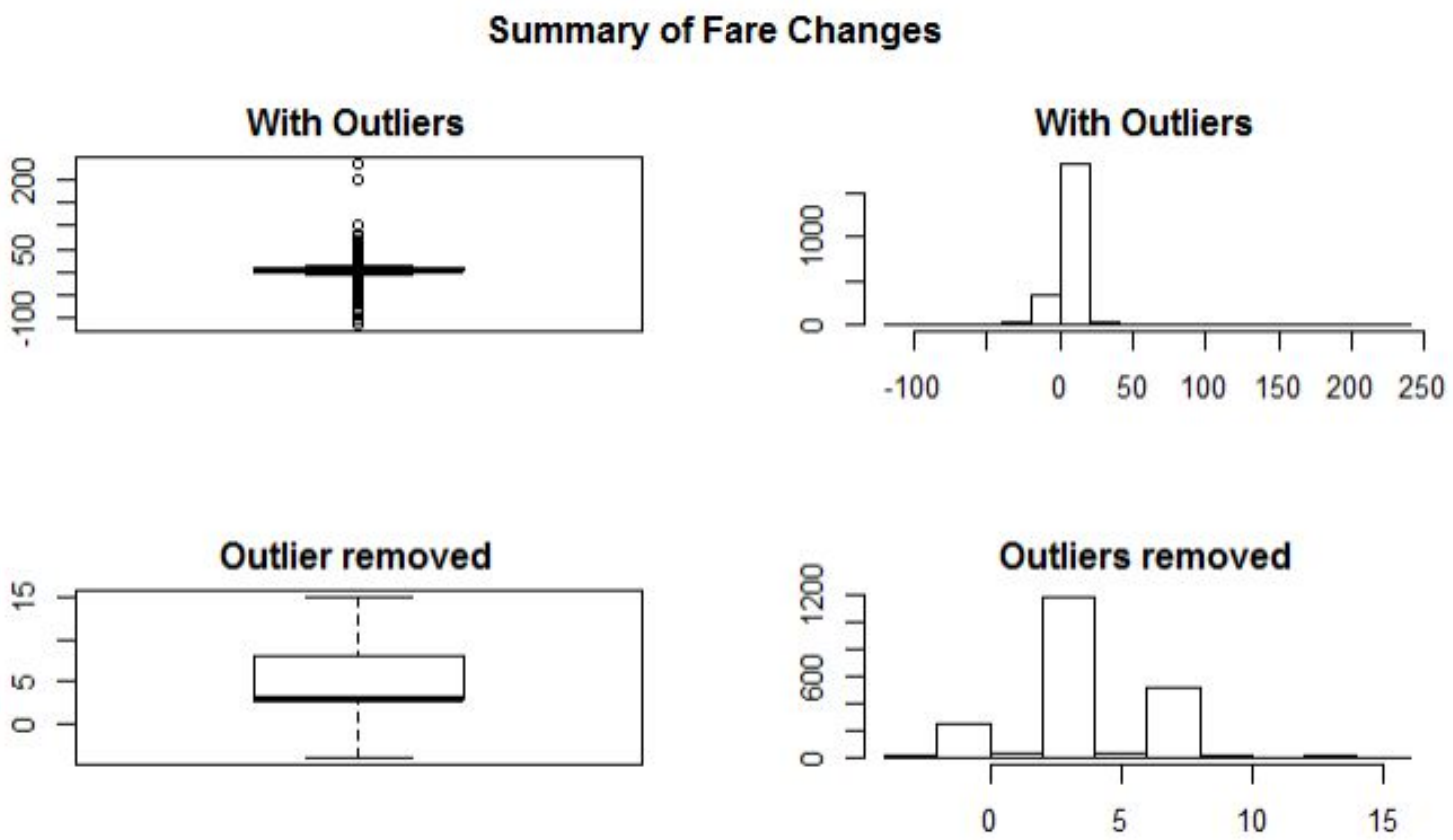
AEROSPACE & DEFENSE | Tue Jan 5, 2016 | 4:31pm EST

# U.S. airlines raise fares $3 in first sector-wide hike since June

## Summary of Fare Changes




Summary of Fare Changes
With Outliers
200
50
-100
With Outliers
1000
0
-100
0
50
100
150
200
250
Outlier removed
15
5
0
Outliers removed
1200
600
0
0
5
10
15

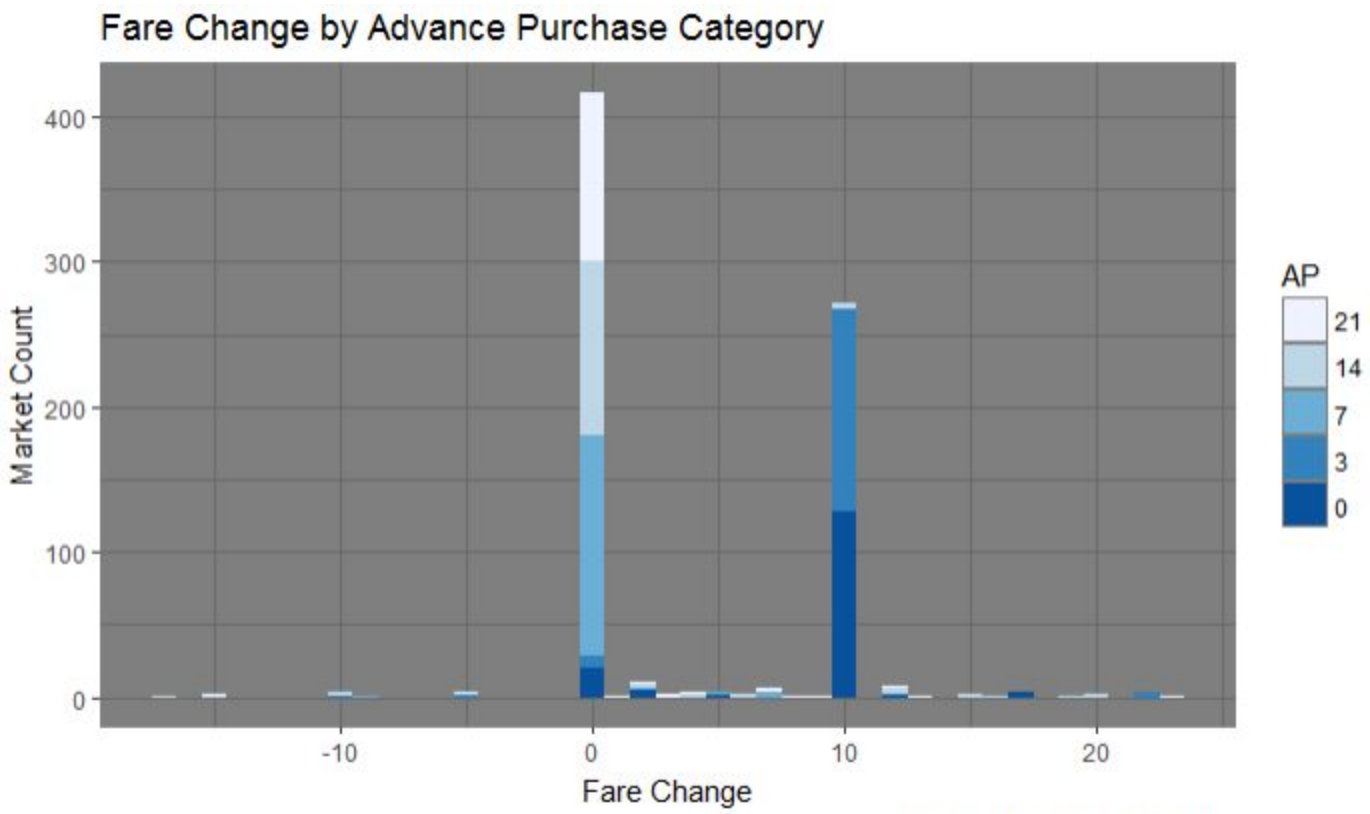
**FORTUNE** | Your Plane Tickets Just Got More Expensive

Fares increased $3 one-way for U.S. domestic flights on American Airlines (AAL, +1.35%), United Continental (UAL, +2.27%), Delta Air Lines (DAL, +1.51%), Southwest Airlines (LUV, +0.42%), and JetBlue Airways (JBLU, +1.97%), each company told Reuters.

# United hikes walkup fares, other majors follow

By Andrew Compart / November 22, 2000

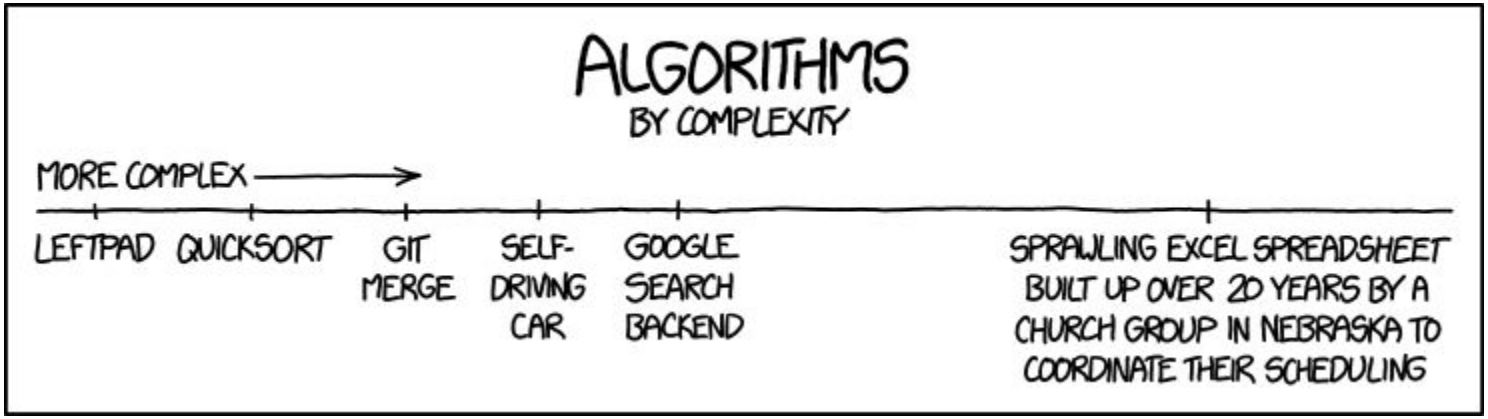
A+ A-




0 and 3 AP are walkup fares

# Spreadsheet Analysis -> R

- Reproducibility of code preserves knowledge
- Less prone to human error
- Processing power is faster -- saves time
- **R scales, spreadsheets don't**



## Departing flights

New York-Kennedy, NY (JFK) to Fort Lauderdale, FL (FLL)

BEST FARE GUARANTEE — jetblue.com has the lowest fares—guaranteed* *Terms apply

| Prev | Sun Jul 2 $142 | Mon Jul 3 $142 | Tue Jul 4 from $114 | Wed Jul 5 $142 | Thu Jul 6 $142 | Next |
|---|---|---|---|---|---|---|

Non-Refundable | Refundable

Compare fare options

Prices displayed are one-way per person including taxes and fees. Additional baggage fees may apply. The order of flights displayed is not neutral with respect to carrier identity; lowest priced JetBlue flights are displayed first, followed by codeshare and interline partner-operated flights, prioritized by price and schedule.
Did you know you can click on a flight number to see its on-time performance?

| Departs | Arrives | Duration | BLUE (0) | BLUE PLUS (1) | BLUE FLEX (2) |
|---|---|---|---|---|---|
| 6:00 AM JFK | 8:55 AM FLL | 2h 55m | $132 — 2 seats left at this price | $155 — 2 seats left at this price | $247 — 2 seats left at this price |
| 301 A320 View seats Amenities | | | | | |
| 8:03 AM JFK | 11:13 AM FLL | 3h 10m | $247 | $270 | $362 |
| 201 A320 View seats Amenities | | | | | |
| 11:06 AM JFK | 2:07 PM FLL | 3h 1m | $142 — 1 seat left at this price | $165 — 1 seat left at this price | $257 — 1 seat left at this price |
| 1 A321 View seats Amenities | | | | | |
| 2:54 PM JFK | 6:07 PM FLL | 3h 13m | $178 | $201 | $293 |
| 1801 A321 View seats Amenities | | | | | |
| 6:42 PM JFK | 9:59 PM FLL | 3h 17m | $142 — 4 seats left at this price | $165 — 4 seats left at this price | $257 — 4 seats left at this price |
| 1401 A321 View seats Amenities | | | | | |
| 8:45 PM JFK | 11:59 PM FLL | 3h 14m | $114 | $137 | $229 |
| 1201 A321 View seats Amenities | | | | | |

- Domain expertise is important
- Context helps you know your data and how to approach it

- **Cautionary note:** *not all NAs are created equally*
- Differentiating missing vs extreme values

- Think holistically about your data
- Don’t automatically NA outliers and extreme values
- **Consider why you might have missing and extreme values**

# Scalable data-driven analysis incorporates strategy and implementation

website %>%
data warehouse %>%
sql/r studio %>%
rmd %>%
report %>%
*strategy/implementation*

- Plan ahead for how your analysis factors into implementation
- Deploy new strategy/model quickly and efficiently
- Ensure decision making data pipeline is reliable and reasonably easy to understand and maintain

# Getting started?

- Start by automating workflows
- Reproducibility - time saved is always an easy sell

Keep a list of common analyses/processes

- Daily/weekly/monthly reports
- Template write ups
- Data processing

Once you've built the groundwork for data processing in R, easier to build out models and more sophisticated analysis from there.

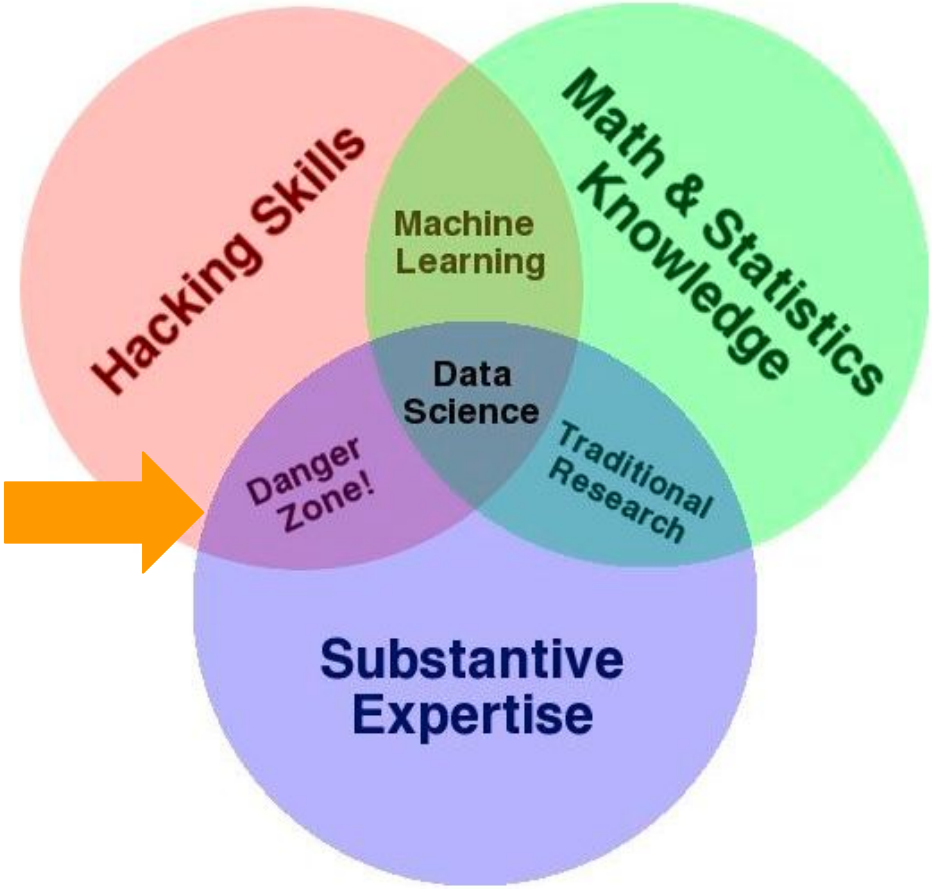
# R for Everyone

- Base R is hard to teach
  - But not everyone needs to gain fluency to start using R
  - Applied approach lowers the barrier of entry to R programming

- Data processing to start
  - Differentiating data types
  - R Notebooks for training/prototyping
  - Common workflows: dplyr, tidyr, lubridate

- Excel and SQL as a frame of reference
  - For those with limited programming experience
  - Excel vlookups -> joins
  - Sorting, filtering, aggregating

# Danger! Not everyone should be fitting models!



# R makes the world go ‘round -- really!

- R can serve wide range of users involved in decision making process
- Building data literacy across workgroups is crucial for scalable analysis

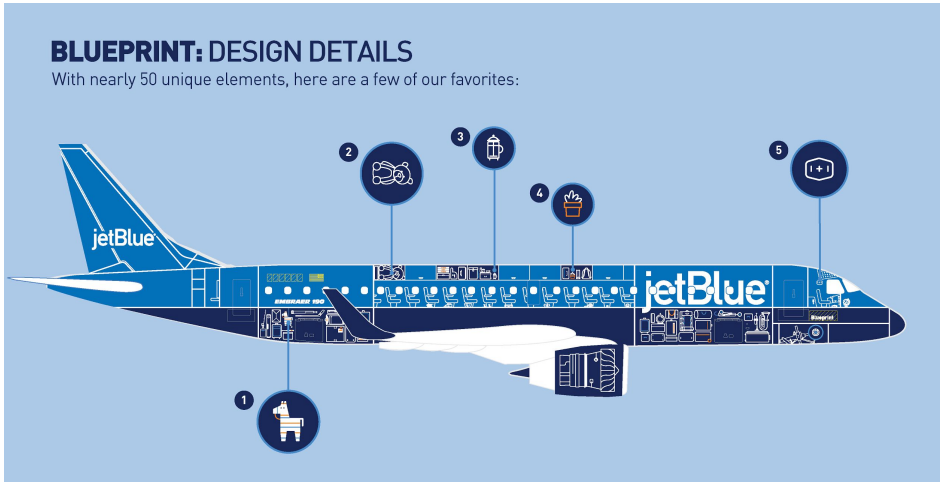
- Beyond data wrangling:
  - Summary statistics
  - Data visualization: ggplot2
  - Automating reports

- Level up!
  - Interactive data viz
  - Dashboards
  - Heuristics
  - Decision trees

# Also an airline nerd?

nycflights13 package
- github.com/hadley/nycflights13

Airline Data Reported Publicly
- Bureau of Transportation Statistics
- transtats.bts.gov



Blue®

jetBlue®

Thank you!