A photograph of a baseball game in progress. In the foreground, a yellow padded fence runs diagonally across the frame. A baseball player's arm, wearing a dark jersey sleeve and a baseball glove, is extended over the fence, reaching for a ball. The background shows a large crowd of spectators seated in the stadium stands under bright daylight. The infield dirt and a portion of the green outfield grass are visible.

Home Run Probability Based on Hit Distance and Direction

Tess Kolp



BACKGROUND

- Hit probability was introduced by Statcast in 2017
- Attempts to answer the question “Based on the **exit velocity** and **launch angle** of the batted ball, how likely was the ball to land for a hit?”



QUESTIONS

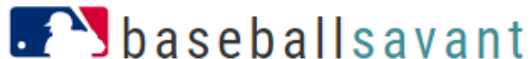
- Based on the distance and spray direction of a batted ball, what is the likelihood of it being a home run?
- Who hit long fly outs? (Who was “unlucky”?)
- Who hit home runs that just barely got out? (Who was “lucky”?)
- Where do these long fly outs and short home runs occur?



MODEL SELECTION

- Using the logistic regression model
- A home run is considered a “success” (1)
- Anything else is considered a “failure” (0)

COLLECTING THE DATA

[About](#)[Daily Matchups](#)[Statcast Leaderboards](#)[Statcast Search](#)[Applications ▾](#)

Statcast Search

An application that lets you search MLB.com's Statcast database. It includes metrics such as Perceived Velocity (PV), Spin Rate (SR), Exit Velocity (EV), Hit Distance (DST), Launch Angle (LA), Batted Ball Direction (BBD), xBA, xSLG, xwOBA, and Quality of Contact.

* Some of these queries are very complicated and take time to run. If the page doesn't fully load, refresh the page and try again.

* Note: Pitch velocities from 2008-16 are via Pitch F/X, and adjusted to roughly out-of-hand release point. All velocities from 2017 and beyond are Statcast, which are reported out-of-hand.

* Note: For the limited subset of batted balls not tracked directly, estimates are included based on the process described [here](#).

Pitch Type:	<input type="text"/>	PA Result:	<input type="text"/>	Batted Ball Type:	<input type="text"/>
Pitch Result:	<input type="text"/>	Gameday Zones:	<input type="text"/>	Venue:	<input type="text"/>
Batted Ball Location:	<input type="text"/>	Detailed Zones:	<input type="text"/>	Season Type:	<input type="text" value="Regular Season"/>
Count:	<input type="text"/>	Season:	<input type="text" value="2017"/>	Situation:	<input type="text"/>
Player Type:	<input type="text" value="Pitcher"/>	Outs:	<input type="text"/>	Opponent:	<input type="text"/>
Pitcher Handedness:	<input type="text"/>	Batter Handedness:	<input type="text"/>	Quality of Contact:	<input type="text"/>
Game Date >=	<input type="text"/>	Game Date <=	<input type="text"/>		
Team:	<input type="text"/>	Position:	<input type="text"/>		
Runners On:	<input type="text"/>	Home or Away:	<input type="text"/>	Batters:	<input type="text" value="Enter Player Name..."/>
Flags:	<input type="text"/>			Pitchers:	<input type="text" value="Enter Player Name..."/>
Metric Range:	<input type="text"/>				
Inning:	<input type="text"/>	Min # of Total Pitches:	<input type="text" value="None"/>	Min # of Results:	<input type="text" value="None"/>
Group By:	<input type="text" value="Player Name"/>	Sort By:	<input type="text" value="Pitches"/>	Sort Order:	<input type="text" value="Desc"/>
<input type="checkbox"/> Only Show Plays with Video		Min ABs:	<input type="text" value="None"/>		

DATA COLLECTED

- All batted balls hit greater than 325 feet in 2017 (19,230 data points)

	A	B	C	D	E	F	G	H
1	player_name	events	home_team	hc_x	hc_y	hit_distance_sc	launch_speed	launch_angle
2	Stephen Vogt	field_out	MIN	56.53	84.21	325	96	40.187
3	Michael Brantley	field_error	CLE	56.89	80.06	325	94.2	28.279
4	Kyle Seager	sac_fly	SEA	201.18	83.01	325	97.7	23.263
5	Justin Upton	double	LAA	96.86	64.84	325	90.5	21.191
6	Adrian Gonzalez	field_out	STL	76.69	74.99	325	91.3	20.763
7	Kevin Pillar	field_out	TOR	148.08	63.55	325	92.6	37.228
8	Matt Adams	field_out	ATL	194.87	79.2	325	110.7	17.956
9	Bryce Harper	field_out	WSH	133.81	61.77	325	111	12.13
10	Xander Bogaerts	field_out	BOS	125.83	78.57	325	86.2	30.85
11	Cameron Maybin	field out	SEA	209.52	88.76	325	96.9	23.524

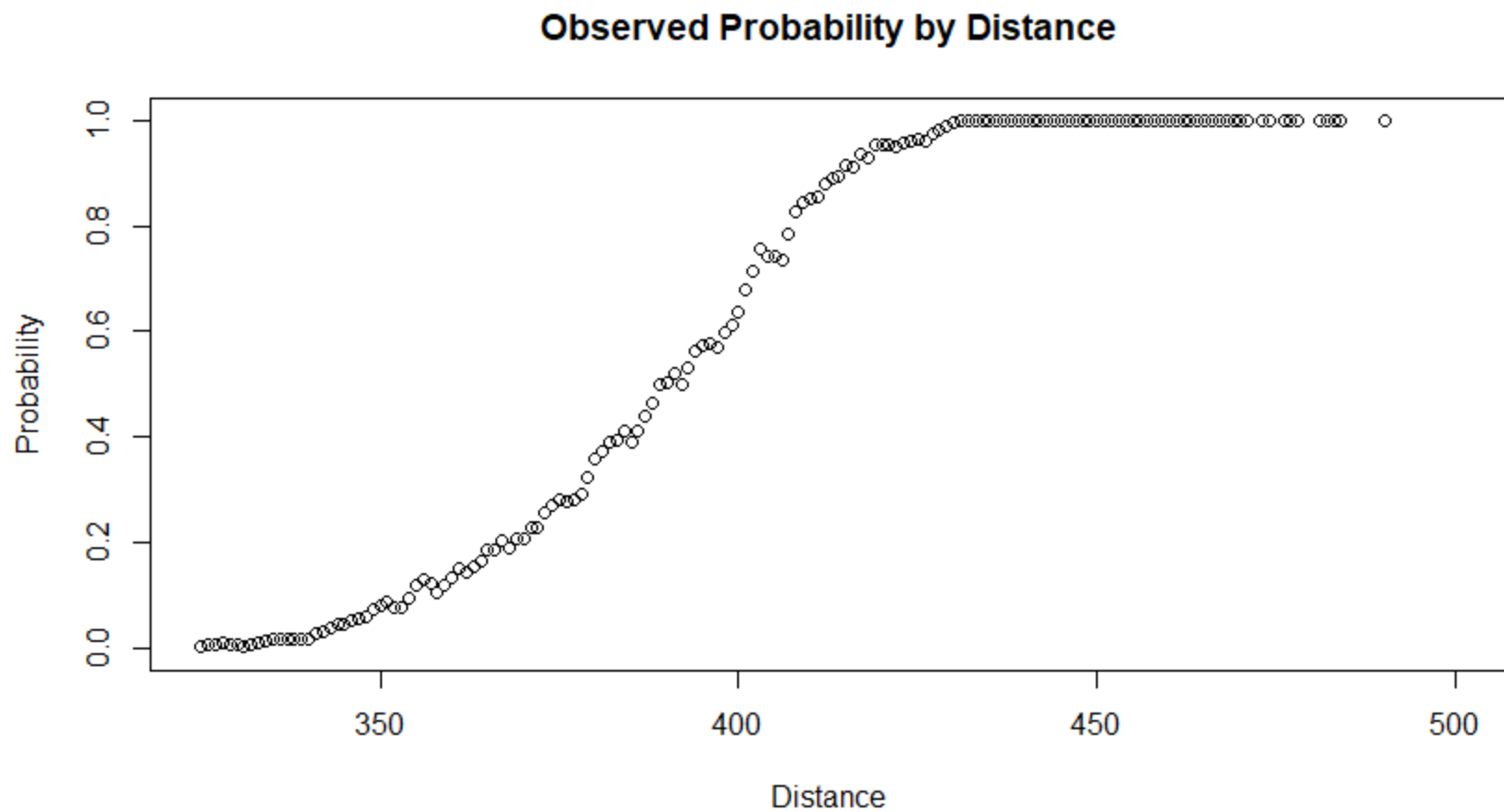


CREATING MY OWN MODEL



DISTANCE ONLY MODEL

INITIAL OBSERVATIONS



GLM OUTPUT

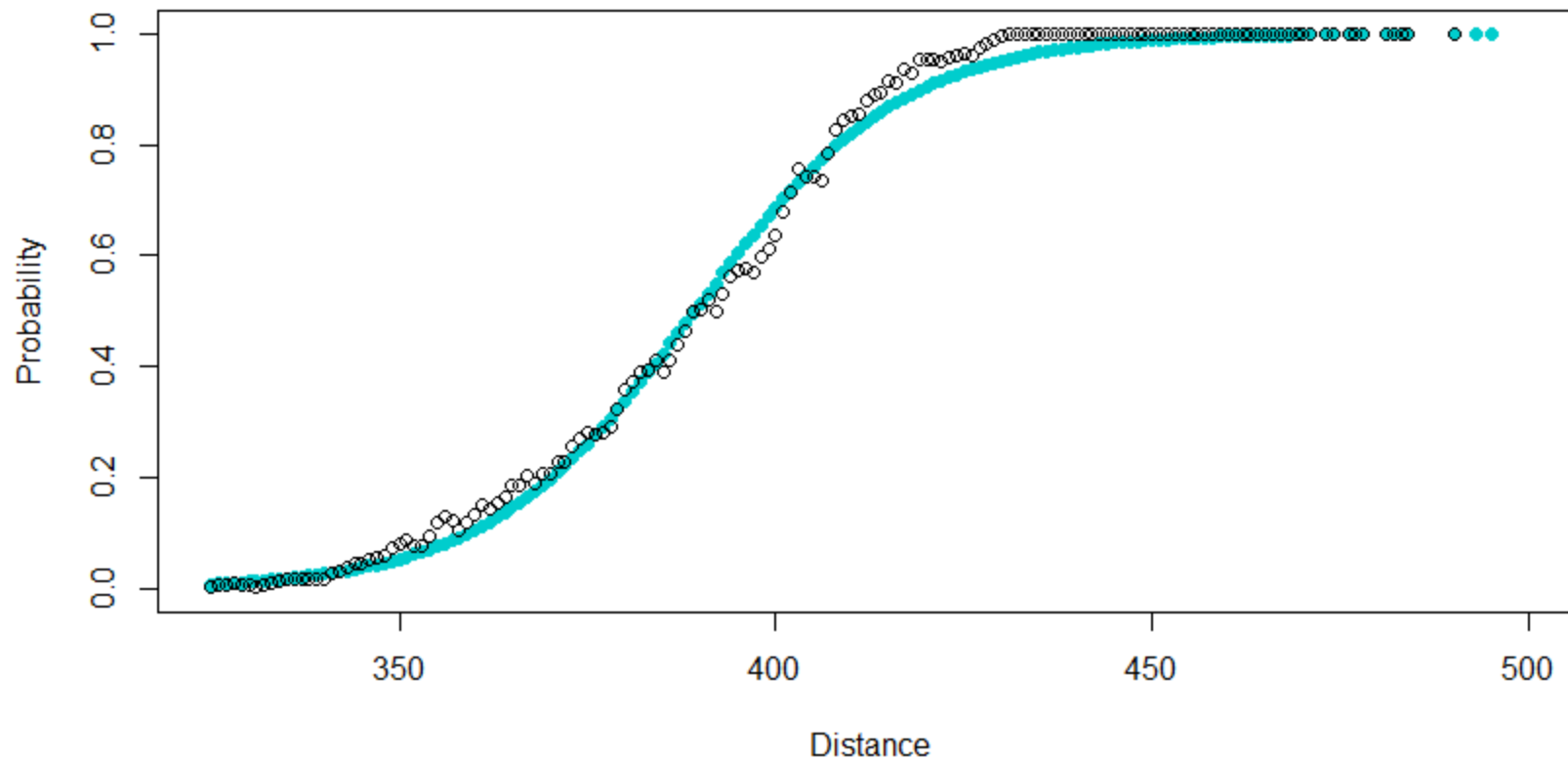
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-28.301748	0.403003	-70.23	<2e-16 ***
distance	0.072716	0.001055	68.93	<2e-16 ***

$$P(hr ; d) = \frac{e^{-28.30+0.07d}}{1 + e^{-28.30+0.07d}}$$

COMPARING MODEL TO OBSERVED PROBABILITIES

Observed and Predicted Probabilities by Distance

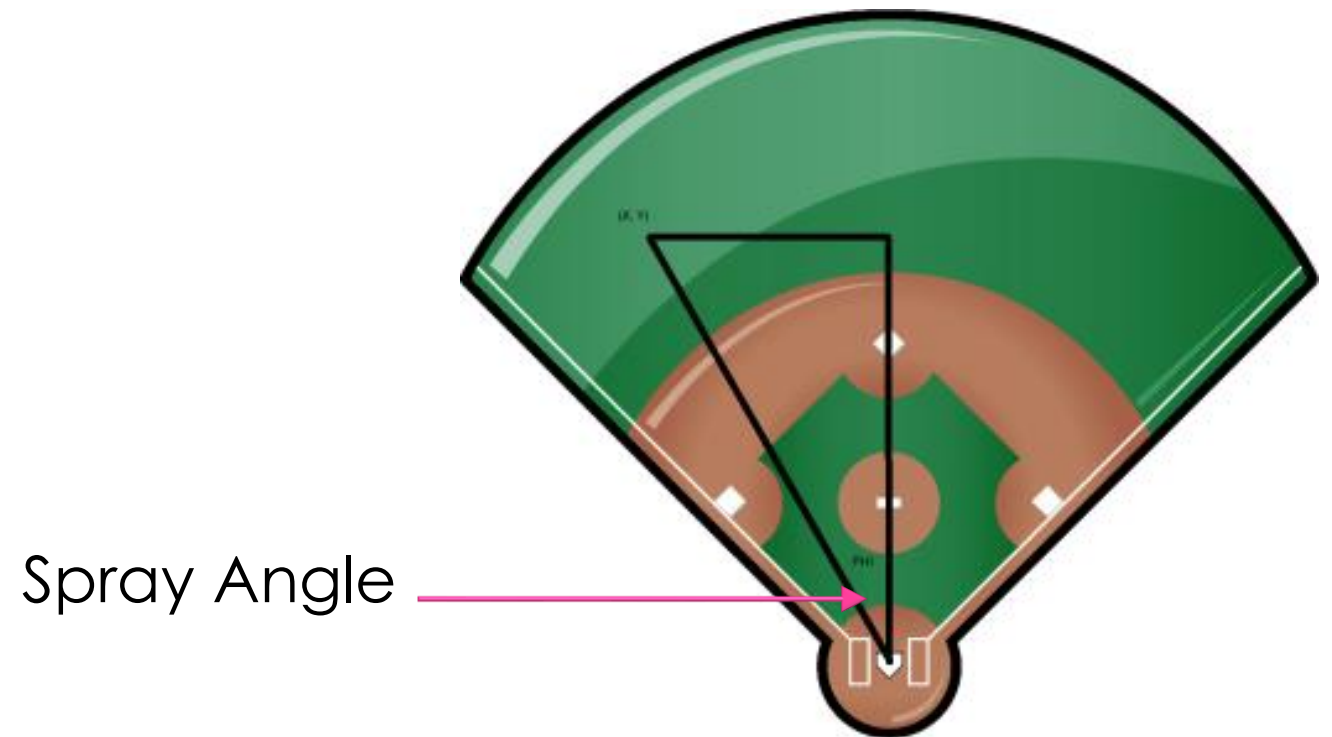




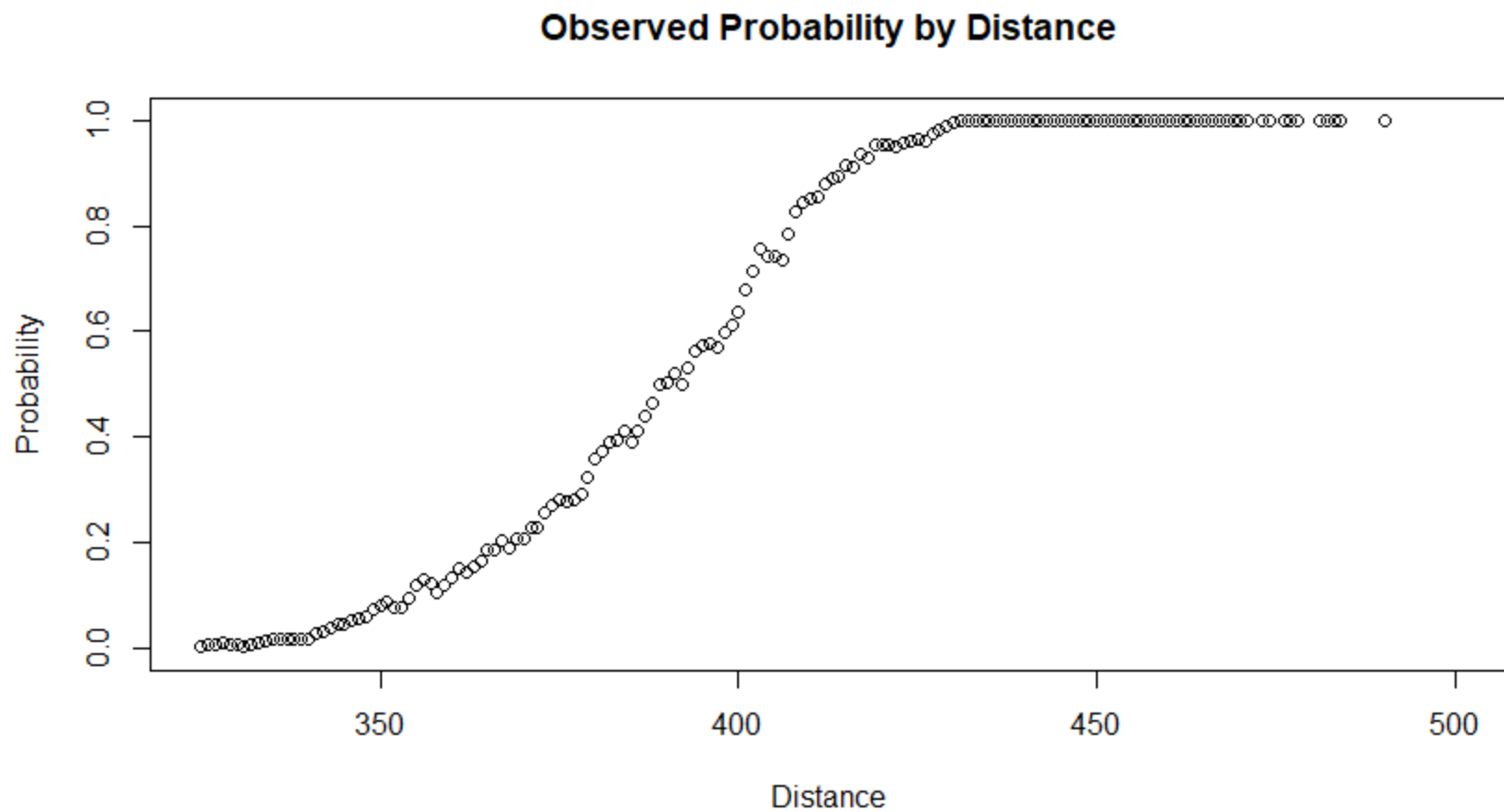
DISTANCE AND DIRECTION MODEL

GETTING ANGLE

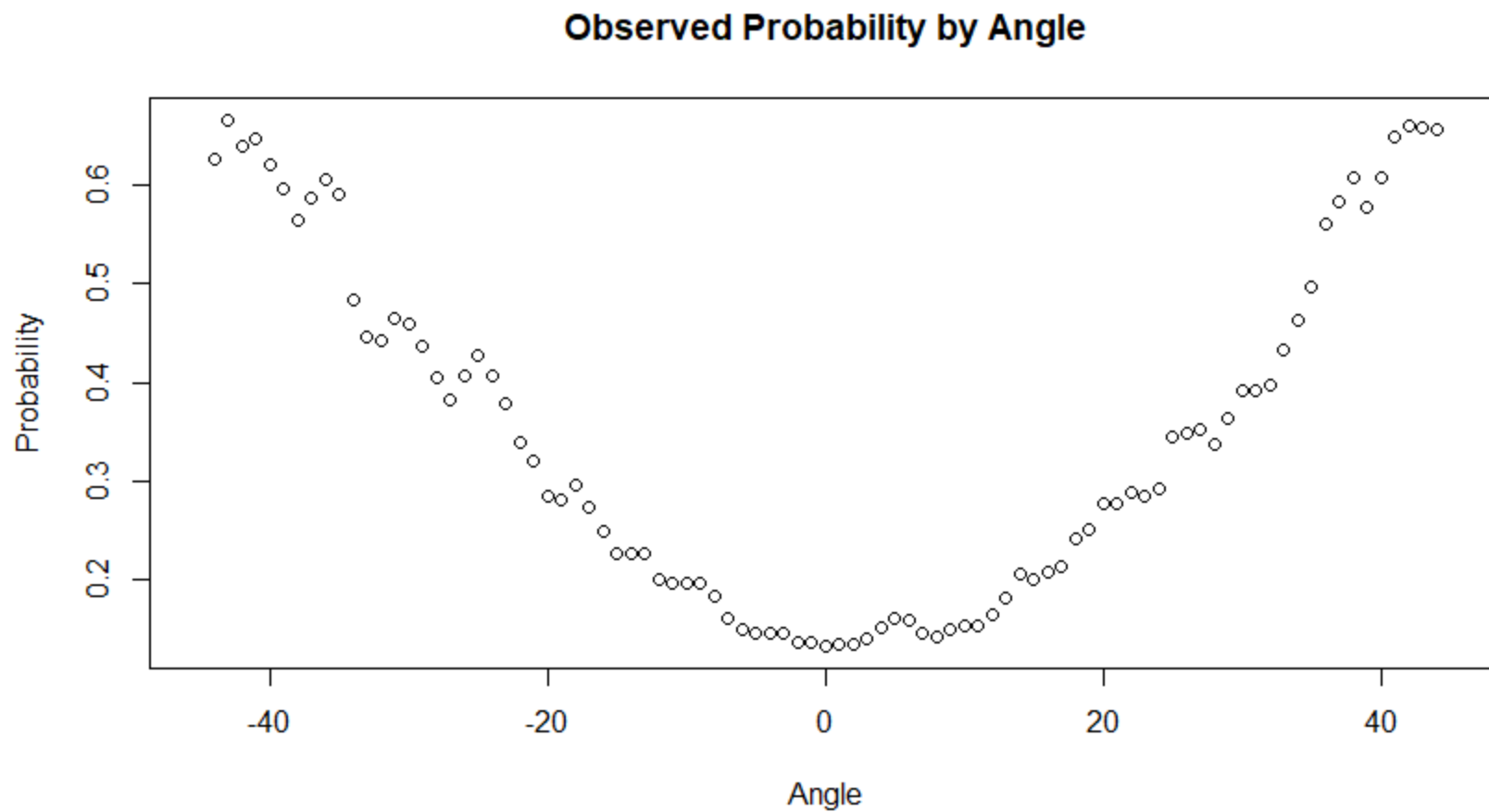
- Transform hc_x and hc_y values so that home plate is $(x=0,y=0)$
- Convert new x and y coordinates into degrees with arctan



INITIAL OBSERVATIONS



INITIAL OBSERVATIONS



CHECKING ASSUMPTIONS

- **Assumption 1: Observations are independent**
 - Plate appearances are independent



CHECKING ASSUMPTIONS

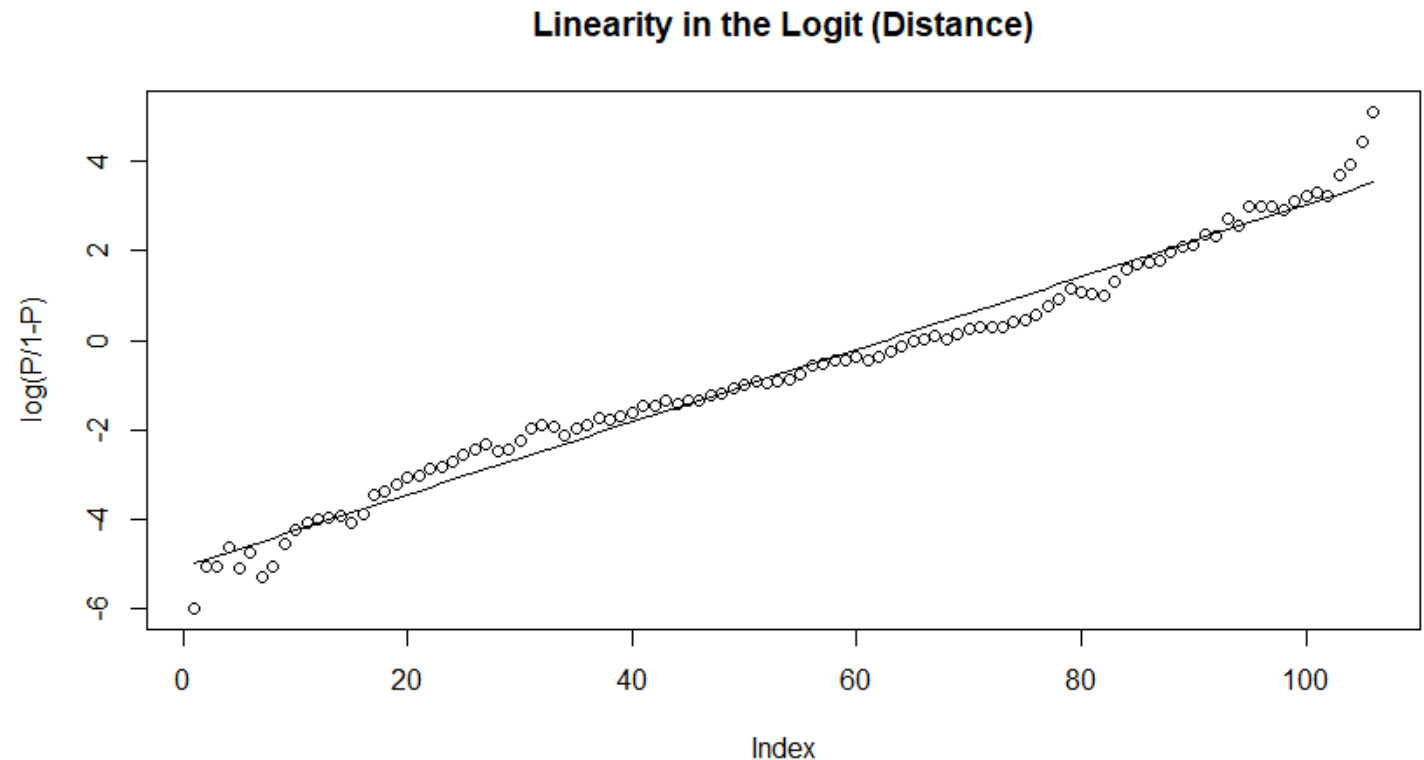
- **Assumption 2: Lack of strongly influential outliers**
 - 19 inside the park home runs left in (short home runs)
 - 8 non home runs 426+ ft. (long non home runs)
 - 4 in Detroit, 1 in Colorado
 - About 0.1% of data
 - Overall distances range from 325 ft. to 495 ft.
- Note: 87 points were removed for having an angle >45 or <-45



CHECKING ASSUMPTIONS

- **Assumption 3: linearity in the logit**

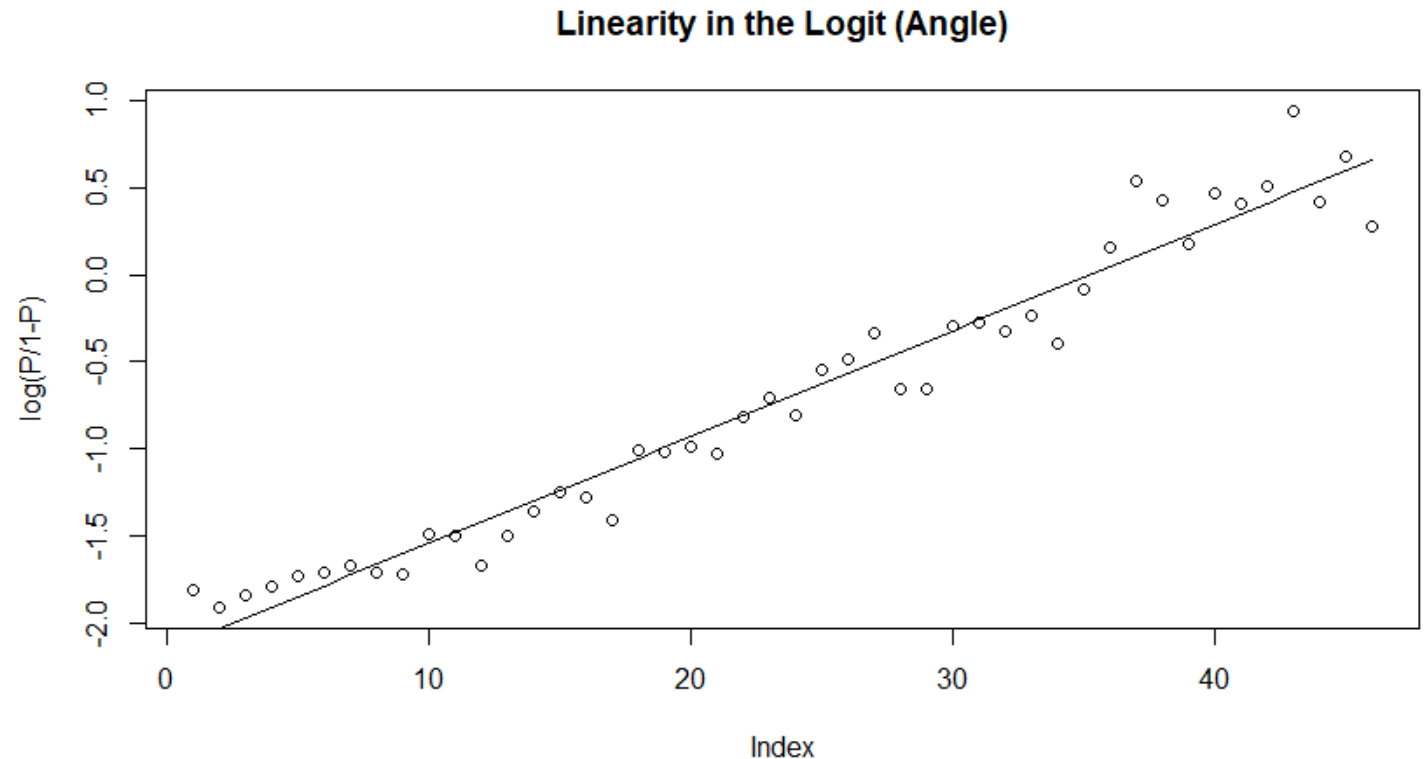
As distance increases, the probability of hitting a home run increases



CHECKING ASSUMPTIONS

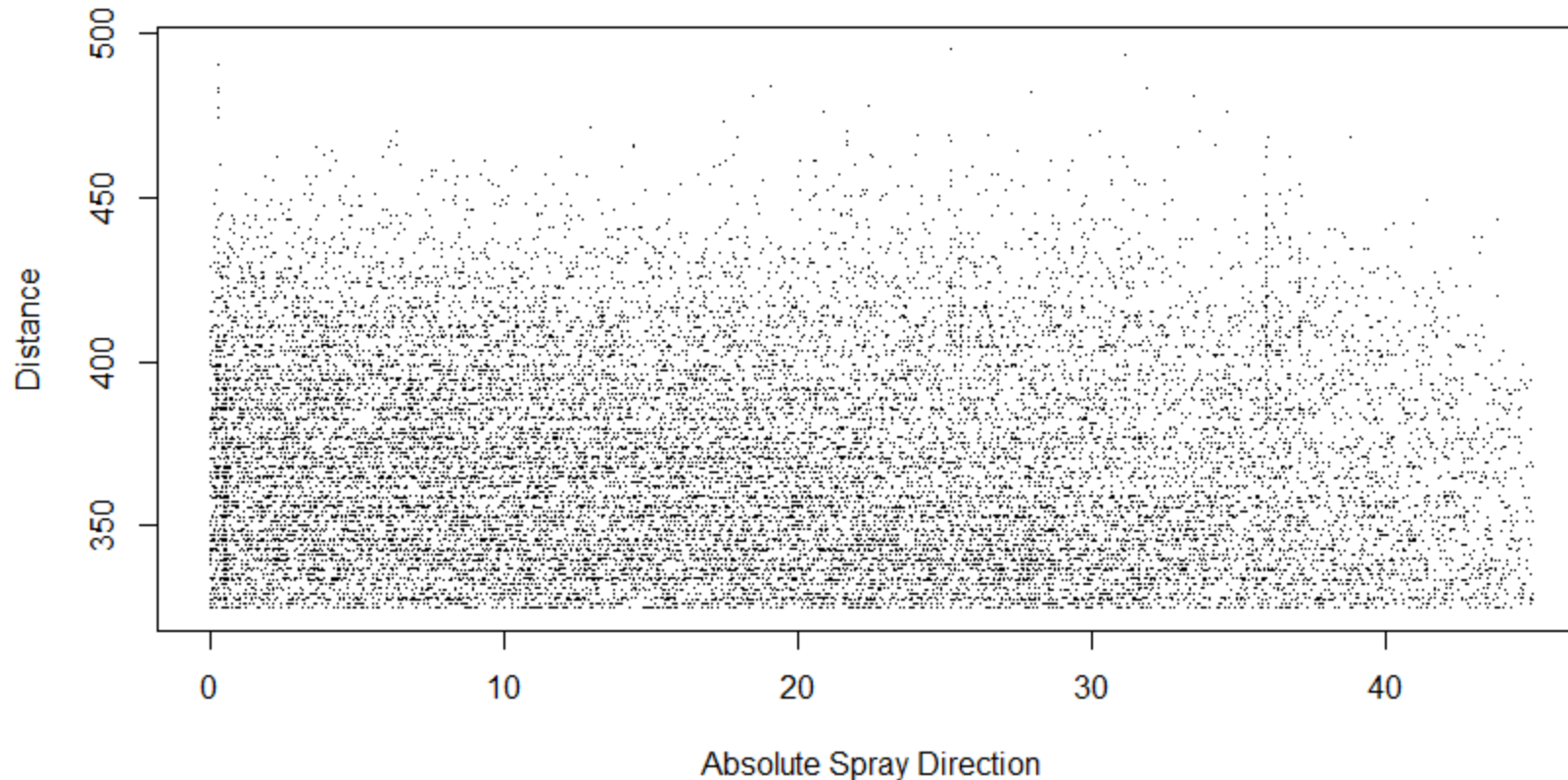
- **Assumption 3: linearity in the logit**

As the absolute value of spray angle increases, the probability of hitting a home run increases



CHECKING ASSUMPTIONS

- **Assumption 4: Little to no multicollinearity**



$$R^2=0.006$$



GLM OUTPUT

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-66.304031	1.189777	-55.73	<2e-16 ***
distance	0.158885	0.002870	55.35	<2e-16 ***
ad	0.262412	0.005282	49.68	<2e-16 ***

$$P(hr ; d, a) = \frac{e^{-66.30+0.16d+0.26|a|}}{1 + e^{-66.30+0.16d+0.26|a|}}$$



AIC COMPARISON

Distance only model

$AIC=13,240$

Distance & Direction model

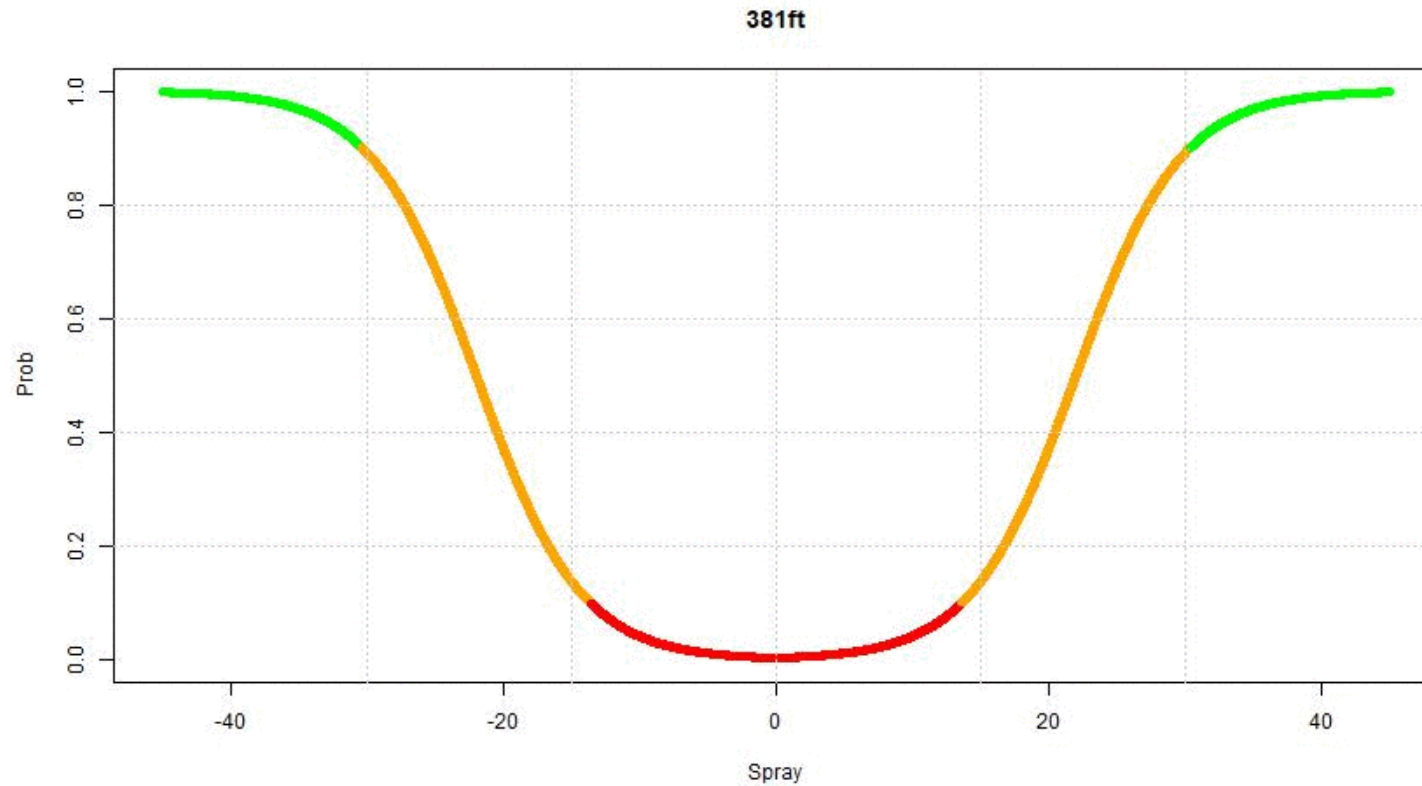
$AIC=6,530$

(Smaller is better)

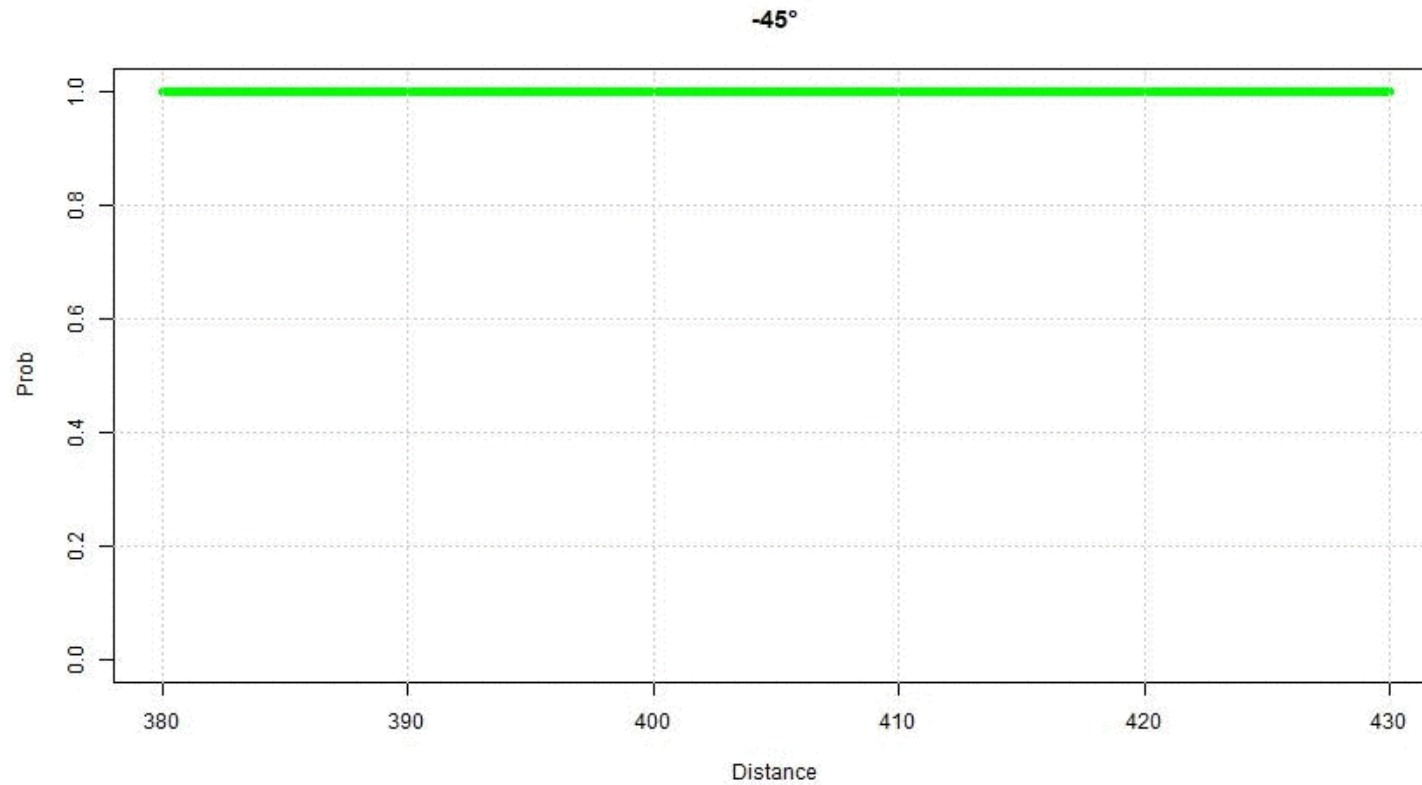
HEAT MAP

	-45	-42	-39	-36	-33	-30	-27	-24	-21	-18	-15	-12	-9	-6	-3	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45
430	1	1	1	1	1	1	1	1	1	1	1	0.99	0.99	0.97	0.94	0.88	0.94	0.97	0.99	0.99	1	1	1	1	1	1	1	1	1	1	1
428	1	1	1	1	1	1	1	1	1	1	1	0.99	0.98	0.96	0.92	0.85	0.92	0.96	0.98	0.99	1	1	1	1	1	1	1	1	1	1	1
426	1	1	1	1	1	1	1	1	1	1	1	0.99	0.98	0.95	0.9	0.8	0.9	0.95	0.98	0.99	1	1	1	1	1	1	1	1	1	1	1
424	1	1	1	1	1	1	1	1	1	1	0.99	0.99	0.97	0.93	0.86	0.74	0.86	0.93	0.97	0.99	0.99	1	1	1	1	1	1	1	1	1	1
422	1	1	1	1	1	1	1	1	1	1	0.99	0.98	0.96	0.91	0.82	0.68	0.82	0.91	0.96	0.98	0.99	1	1	1	1	1	1	1	1	1	1
420	1	1	1	1	1	1	1	1	1	0.99	0.99	0.97	0.94	0.88	0.77	0.61	0.77	0.88	0.94	0.97	0.99	0.99	1	1	1	1	1	1	1	1	1
418	1	1	1	1	1	1	1	1	1	0.99	0.98	0.96	0.92	0.84	0.71	0.53	0.71	0.84	0.92	0.96	0.98	0.99	1	1	1	1	1	1	1	1	1
416	1	1	1	1	1	1	1	1	1	0.99	0.98	0.95	0.9	0.8	0.64	0.45	0.64	0.8	0.9	0.95	0.98	0.99	1	1	1	1	1	1	1	1	1
414	1	1	1	1	1	1	1	1	0.99	0.99	0.97	0.93	0.86	0.74	0.56	0.37	0.56	0.74	0.86	0.93	0.97	0.99	0.99	1	1	1	1	1	1	1	1
412	1	1	1	1	1	1	1	1	0.99	0.98	0.96	0.91	0.82	0.68	0.49	0.3	0.49	0.68	0.82	0.91	0.96	0.98	0.99	1	1	1	1	1	1	1	1
410	1	1	1	1	1	1	1	0.99	0.99	0.97	0.94	0.88	0.77	0.6	0.41	0.24	0.41	0.6	0.77	0.88	0.94	0.97	0.99	0.99	1	1	1	1	1	1	1
408	1	1	1	1	1	1	1	0.99	0.98	0.96	0.92	0.84	0.71	0.52	0.33	0.19	0.33	0.52	0.71	0.84	0.92	0.96	0.98	0.99	1	1	1	1	1	1	1
406	1	1	1	1	1	1	0.99	0.99	0.98	0.95	0.89	0.79	0.64	0.44	0.27	0.14	0.27	0.44	0.64	0.79	0.89	0.95	0.98	0.99	0.99	1	1	1	1	1	1
404	1	1	1	1	1	1	0.99	0.98	0.97	0.93	0.86	0.74	0.56	0.37	0.21	0.11	0.21	0.37	0.56	0.74	0.86	0.93	0.97	0.98	0.99	1	1	1	1	1	1
402	1	1	1	1	1	1	0.99	0.98	0.96	0.91	0.82	0.67	0.48	0.3	0.16	0.08	0.16	0.3	0.48	0.67	0.82	0.91	0.96	0.98	0.99	1	1	1	1	1	1
400	1	1	1	1	1	0.99	0.99	0.97	0.94	0.88	0.77	0.6	0.4	0.24	0.12	0.06	0.12	0.24	0.4	0.6	0.77	0.88	0.94	0.97	0.99	0.99	1	1	1	1	1
398	1	1	1	1	1	0.99	0.98	0.96	0.92	0.84	0.7	0.52	0.33	0.18	0.09	0.04	0.09	0.18	0.33	0.52	0.7	0.84	0.92	0.96	0.98	0.99	1	1	1	1	1
396	1	1	1	1	0.99	0.99	0.98	0.95	0.89	0.79	0.63	0.44	0.26	0.14	0.07	0.03	0.07	0.14	0.26	0.44	0.63	0.79	0.89	0.95	0.98	0.99	0.99	1	1	1	1
394	1	1	1	1	0.99	0.98	0.96	0.91	0.82	0.67	0.48	0.29	0.16	0.08	0.04	0.02	0.04	0.08	0.16	0.29	0.48	0.67	0.82	0.91	0.96	0.98	0.99	1	1	1	1
392	1	1	1	1	0.99	0.98	0.97	0.93	0.86	0.73	0.56	0.36	0.21	0.11	0.05	0.02	0.05	0.11	0.21	0.36	0.56	0.73	0.86	0.93	0.97	0.98	0.99	1	1	1	1
390	1	1	0.99	0.98	0.97	0.93	0.86	0.73	0.55	0.36	0.2	0.1	0.05	0.02	0.01	0.01	0.01	0.02	0.05	0.1	0.2	0.36	0.55	0.73	0.86	0.93	0.97	0.98	0.99	1	1
388	1	1	0.99	0.99	0.98	0.95	0.89	0.79	0.63	0.44	0.26	0.14	0.07	0.03	0.01	0.01	0.01	0.03	0.07	0.14	0.26	0.44	0.63	0.79	0.89	0.95	0.98	0.99	0.99	1	1
386	1	1	1	0.99	0.98	0.96	0.92	0.84	0.7	0.52	0.33	0.18	0.09	0.04	0.02	0.01	0.02	0.04	0.09	0.18	0.33	0.52	0.7	0.84	0.92	0.96	0.98	0.99	1	1	1
384	1	1	1	0.99	0.99	0.97	0.94	0.88	0.76	0.59	0.4	0.23	0.12	0.06	0.03	0.01	0.03	0.06	0.12	0.23	0.4	0.59	0.76	0.88	0.94	0.97	0.99	0.99	1	1	1
382	1	0.99	0.99	0.97	0.94	0.87	0.76	0.59	0.4	0.23	0.12	0.06	0.03	0.01	0.01	0	0.01	0.01	0.03	0.06	0.12	0.23	0.4	0.59	0.76	0.87	0.94	0.97	0.99	0.99	1
380	1	1	0.99	0.98	0.95	0.91	0.81	0.67	0.48	0.29	0.16	0.08	0.04	0.02	0.01	0	0.01	0.02	0.04	0.08	0.16	0.29	0.48	0.67	0.81	0.91	0.95	0.98	0.99	1	1

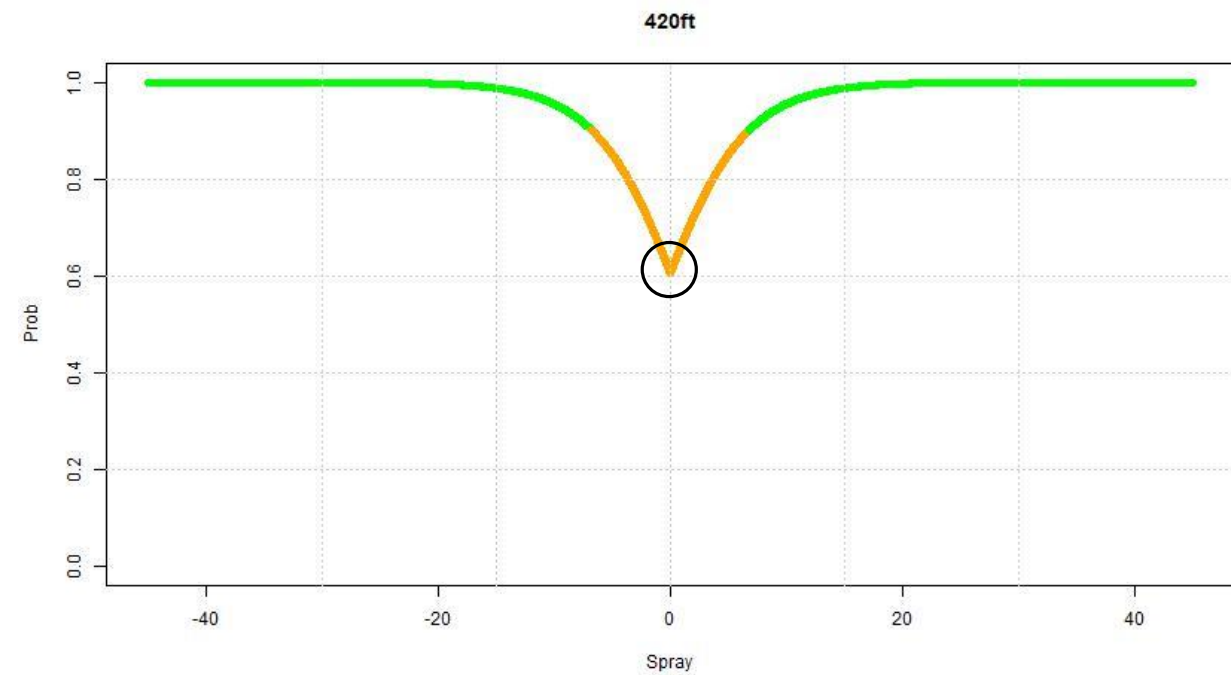
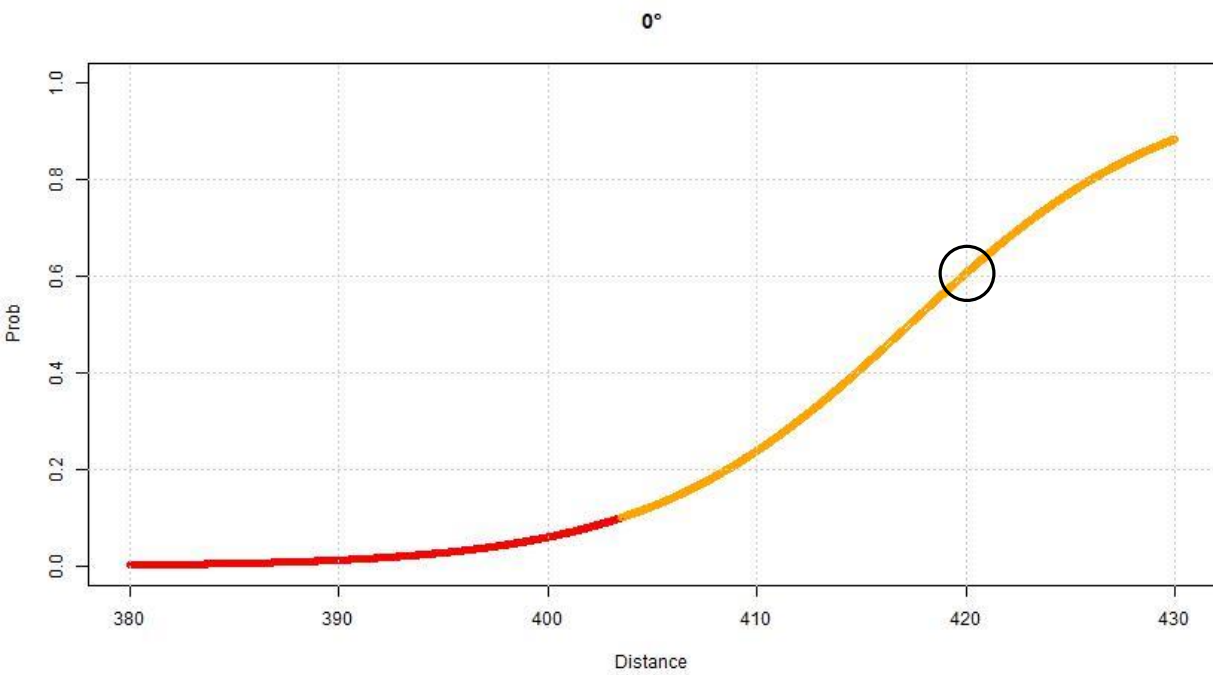
CHANGE IN PROBABILITY BY CHANGE IN DISTANCE



CHANGE IN PROBABILITY BY CHANGE IN DIRECTION



CONSIDER

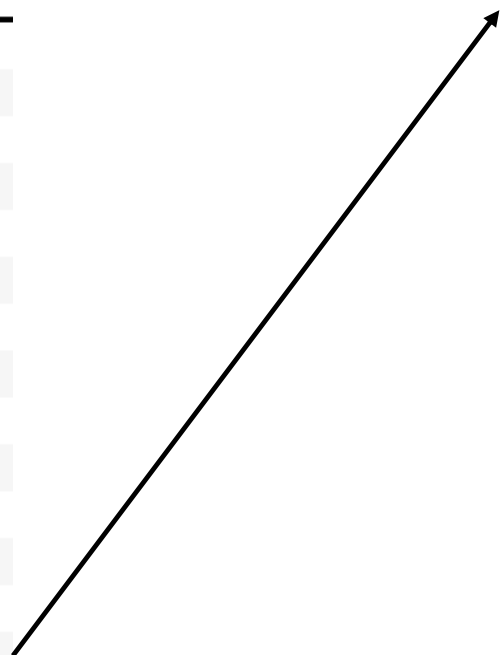




OTHER MODELS

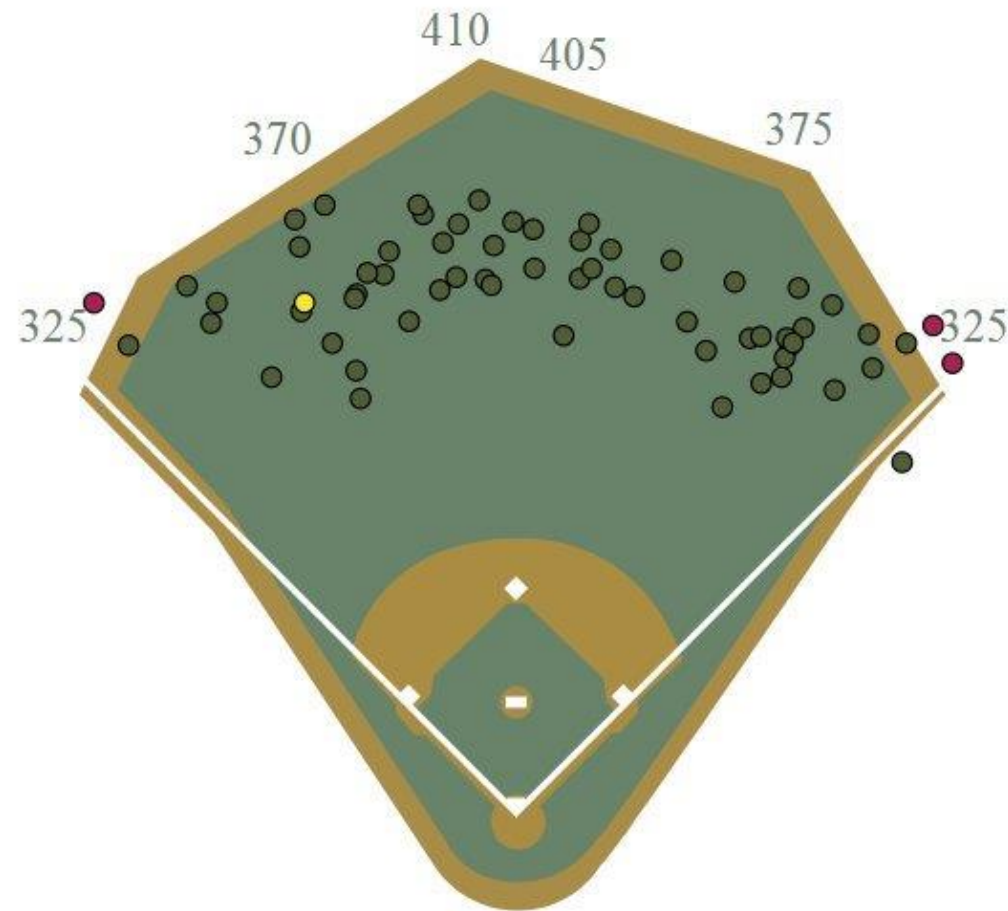
LAUNCH ANGLE & EXIT VELOCITY

Exit Velocity (MPH)	HR %
123	0.0
122	0.0
121	100.0
120	0.0
119	45.5
118	18.2
117	28.6
116	24.4
115	21.1
114	24.1
113	22.2
112	23.6
111	21.2
110	21.4
109	22.2
108	22.1
107	21.8
106	20.7
105	19.2
104	17.3



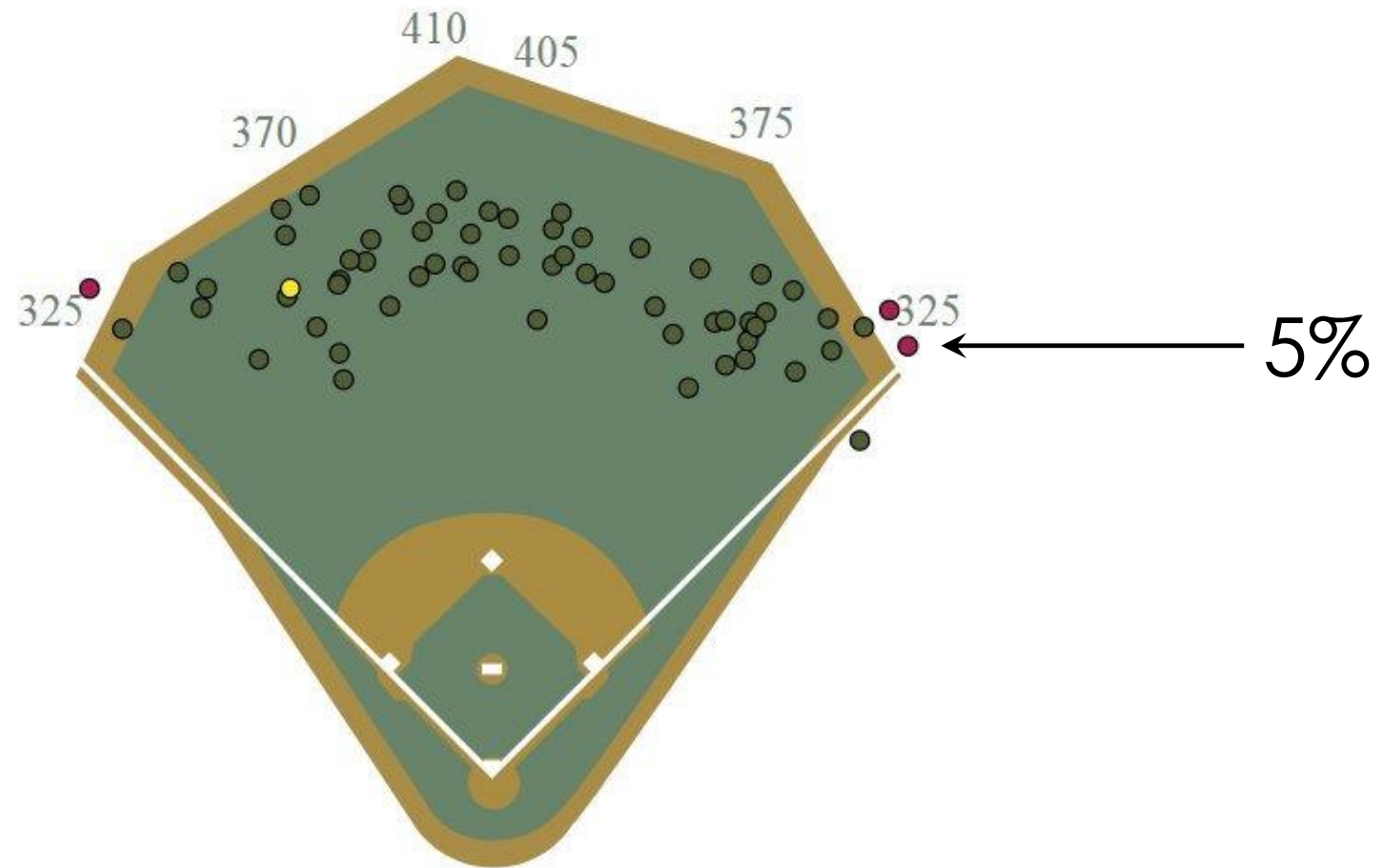
Launch Angle (Deg)	HR %
51	0.0
50	0.0
49	0.0
48	100.0
43	100.0
42	100.0
40	100.0
39	100.0
38	66.7
37	100.0
36	100.0
35	100.0
34	100.0
33	100.0
32	94.7
31	100.0
30	100.0
29	100.0
28	96.7
27	92.9

LAUNCH ANGLE & EXIT VELOCITY



94 MPH, 41°, BA: 0.063 1B%: 0 2B%: 2 3B%: 0 HR %: 5

LAUNCH ANGLE & EXIT VELOCITY



94 MPH, 41°, BA: 0.063 1B%: 0 2B%: 2 3B%: 0 HR %: 5



AUGMENTED HIT PROBABILITY

- Still uses launch angle and exit velocity
- If the ball was hit close to a generic fence line, spray direction was included
- Better estimates higher probabilities
- Note: this model has not been released yet



APPLICATION



THESE TWO FLY OUTS

Jake Lamb

- Launch angle: 26°
- Exit velocity: 100 mph

Matt Adams

- Launch angle: 26°
- Exit velocity: 100 mph



THESE TWO FLY OUTS

Jake Lamb

- Launch angle: 26°
- Exit velocity: 100 mph

46%

Matt Adams

- Launch angle: 26°
- Exit velocity: 100 mph

46%



THESE TWO FLY OUTS

Jake Lamb

- Launch angle: 26°
- Exit velocity: 100 mph
- Distance: 385 feet

46%

Matt Adams

- Launch angle: 26°
- Exit velocity: 100 mph
- Distance: 389 feet

46%

THESE TWO FLY OUTS

Jake Lamb

- Launch angle: 26°
- Exit velocity: 100 mph

46%

- Distance: 385 feet
- **Spray angle: -23°**

Matt Adams

- Launch angle: 26°
- Exit velocity: 100 mph

46%

- Distance: 389 feet
- **Spray angle: -8°**

THESE TWO FLY OUTS

Jake Lamb

- Launch angle: 26°
- Exit velocity: 100 mph

46%

- Distance: 385 feet
- **Spray angle: -23°**

72%

Matt Adams

- Launch angle: 26°
- Exit velocity: 100 mph

46%

- Distance: 389 feet
- **Spray angle: -8°**

8%



THESE TWO HOME RUNS

Chris Young

- Launch angle: 26°
- Exit velocity: 100 mph

46%

Matt Wieters

- Launch angle: 26°
- Exit velocity: 100 mph

46%

THESE TWO HOME RUNS

Chris Young

- Launch angle: 26°
- Exit velocity: 100 mph
- Distance: 399 feet

46%

Matt Wieters

- Launch angle: 26°
- Exit velocity: 100 mph
- Distance: 403 feet

46%

THESE TWO HOME RUNS

Chris Young

- Launch angle: 26°
- Exit velocity: 100 mph

46%

- Distance: 399 feet
- **Spray angle: -32°**

Matt Wieters

- Launch angle: 26°
- Exit velocity: 100 mph

46%

- Distance: 403 feet
- **Spray angle: 3°**

THESE TWO HOME RUNS

Chris Young

- Launch angle: 26°
- Exit velocity: 100 mph

46%

- Distance: 399 feet
- **Spray angle: -32°**

100%

Matt Wieters

- Launch angle: 26°
- Exit velocity: 100 mph

46%

- Distance: 403 feet
- **Spray angle: 3°**

18%

THIS SALVADOR PEREZ FLY OUT

LA & EV model

91% home run
probability

LA=30°

EV=106 mph



My model

97% home run
probability

Distance=426

Angle=8°

THIS MIGUEL CABRERA FLY OUT

LA & EV model
91% home run
probability

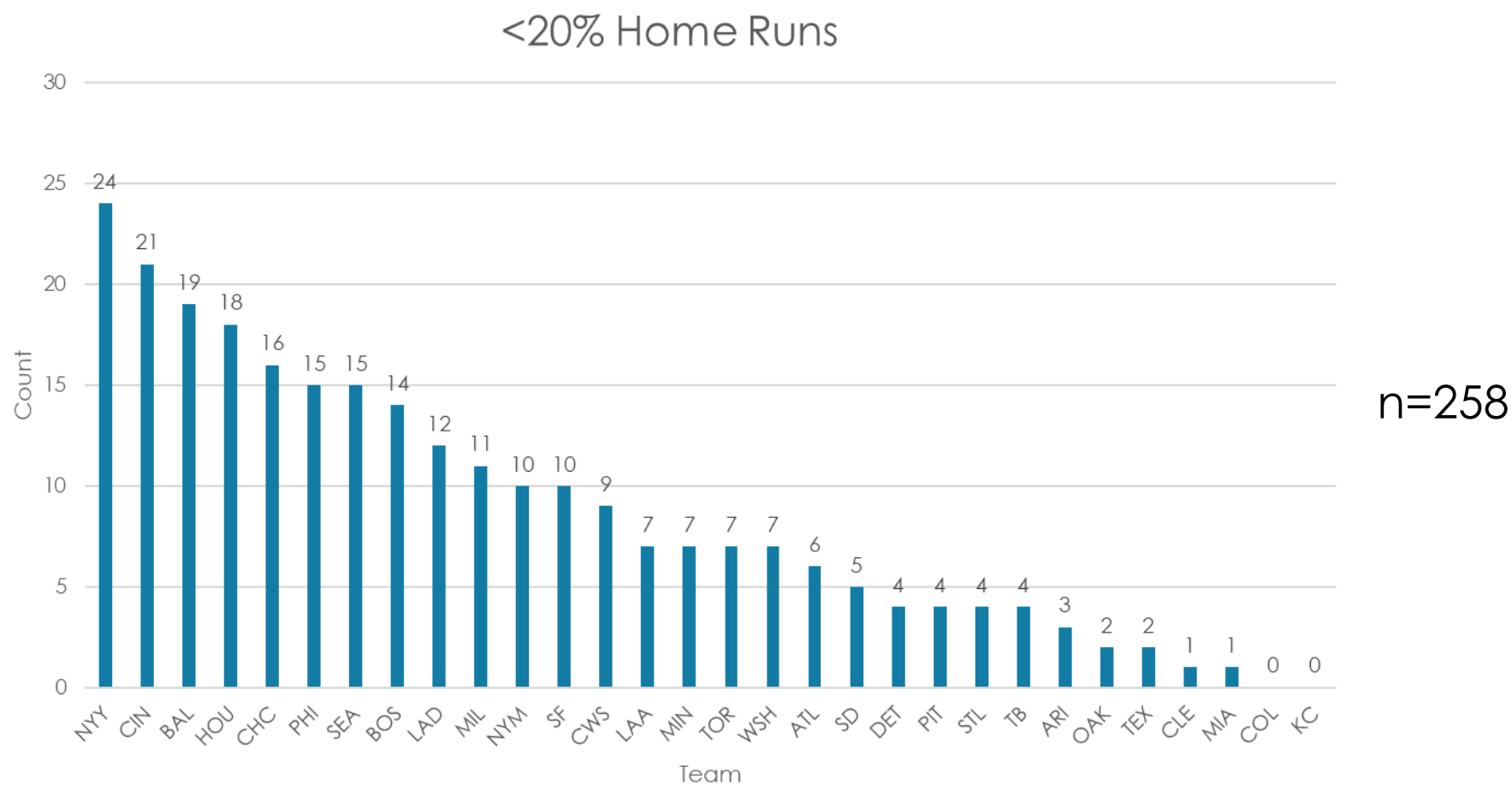
LA=30°
EV=108 mph



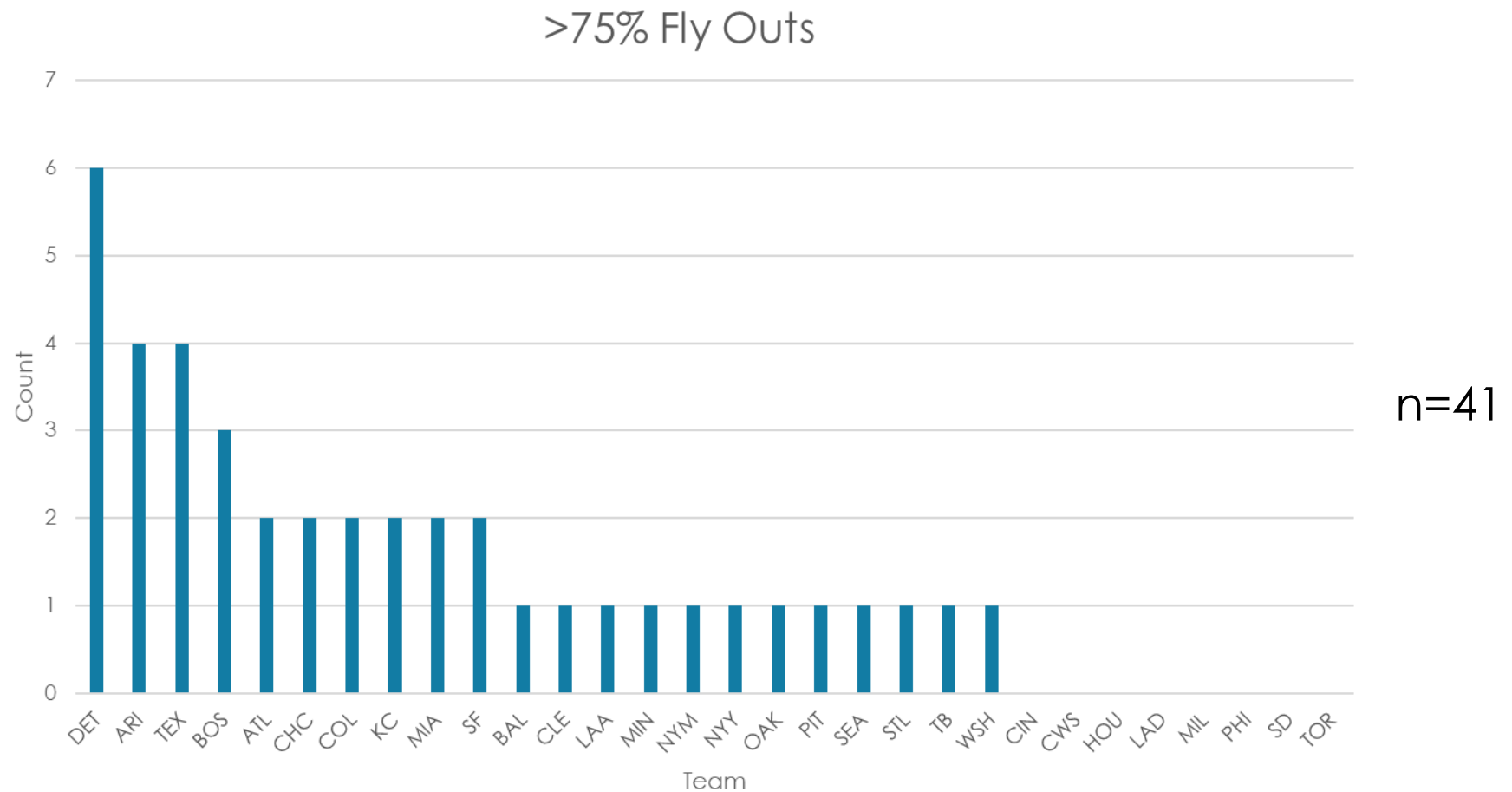
My model
1% home run
probability

Distance=388
Angle=0°

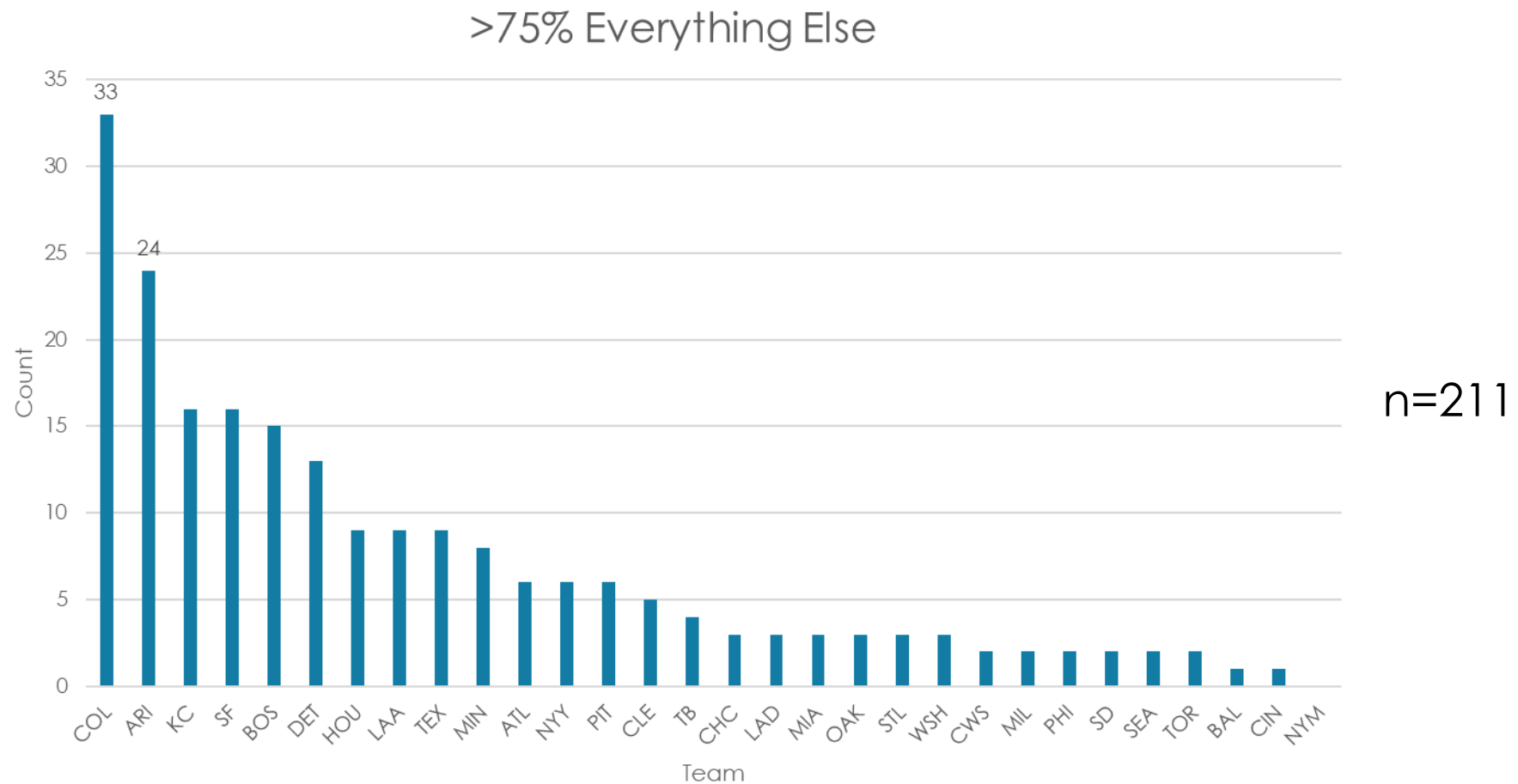
THE “WHERE” QUESTION



THE “WHERE” QUESTION



THE “WHERE” QUESTION





MOVING FORWARD



MOVING FORWARD

- Explaining the sharp dip in probability around 0°
- Finding better ways to compare observed and predicted probabilities
- Keep improving!



THANK YOU!

- Committee members
 - Dr. Steve Ziebarth
 - Dr. Joe McKean
 - Tom Tango
- Family
- Everyone who came today!



QUESTIONS?