

DATA ANALYTICS: BASEBALL PITCHES

Emil Matti

BASEBALL PITCHING GENERAL KNOWLEDGE

- A pitcher stands from 60 ft. 6 inches away from the batter
- Goal is to get the batter out by either throwing three strikes or having the batter hit the ball to one of the fielders
- Pitcher can throw the ball in a way that slows the ball down, speeds the ball up, or give different movement (makes it hard for the hitter to hit the ball)
- <https://www.youtube.com/watch?v=RoGZV6YorFA>

DATASET (KAGGLE)

722,161 pitches thrown from 2015 to 2018

SQL to clean.

- Start speed, end speed, break angle, break length, pitch type, hitter's team score, ball count, strike count, outs, pitch number

In[1]:=

```
mlb=SemanticImport["/Users/emilmatti/Desktop/pitchtype.csv"]
```

Out[1]:=

START_SPEED	END_SPEED	BREAK_ANGLE	BREAK_LENGTH	PITCH_TYPE	B_SCORE	B_COUNT	S_COUNT
79.9	72.4	13.2	14.4	KC	1	0	0
91.6	84.4	33.6	6.0	SI	1	0	0
80.2	74.8	15.6	12.0	KC	1	1	0
80.6	74.1	12.0	13.2	KC	1	1	0
91.1	83.5	30.0	7.2	SI	1	2	0
82.6	76.4	16.8	10.8	CH	1	0	0
90.4	81.8	28.8	6.0	FF	2	0	0
91.5	82.8	30.0	4.8	FF	2	0	1
80.4	73.8	4.8	14.4	KC	2	0	2
91.2	83.3	36.0	6.0	SI	2	0	2
80.1	72.7	16.8	12.0	KC	2	0	0
78.0	71.7	15.6	13.2	KC	2	1	0
91.9	83.7	24.0	4.8	FF	2	1	0
90.9	83.0	28.8	4.8	FF	2	2	0
72.4	65.3	12.0	16.8	CU	0	0	0
88.1	79.9	31.2	4.8	FF	0	1	0
84.0	75.9	2.4	6.0	SL	0	2	0
88.0	79.7	31.2	4.8	FF	0	3	0
85.4	79.0	6.0	6.0	SL	0	3	0
87.4	79.3	27.6	4.8	FF	0	0	0

⤵ ⤶ rows 1–20 of 722161 ⤷ ⤸ ⤹ ⤺ columns 1–10 of 13 ⤻ ⤻



PITCH TYPES (*PITCH_TYPE*)

FF = *Four-seam Fastball*

SL = *Slider*

CH=*Changeup*

CU = *Curveball*

FT = *Two-Seam Fastball*

FT = *Two-Seam Fastball*

SI = *Sinker*

FC = *Cutter*

KC = *Knuckle Curve*

FS = *Splitter*

KN = *Knuckleball*

EP = *Eephus*

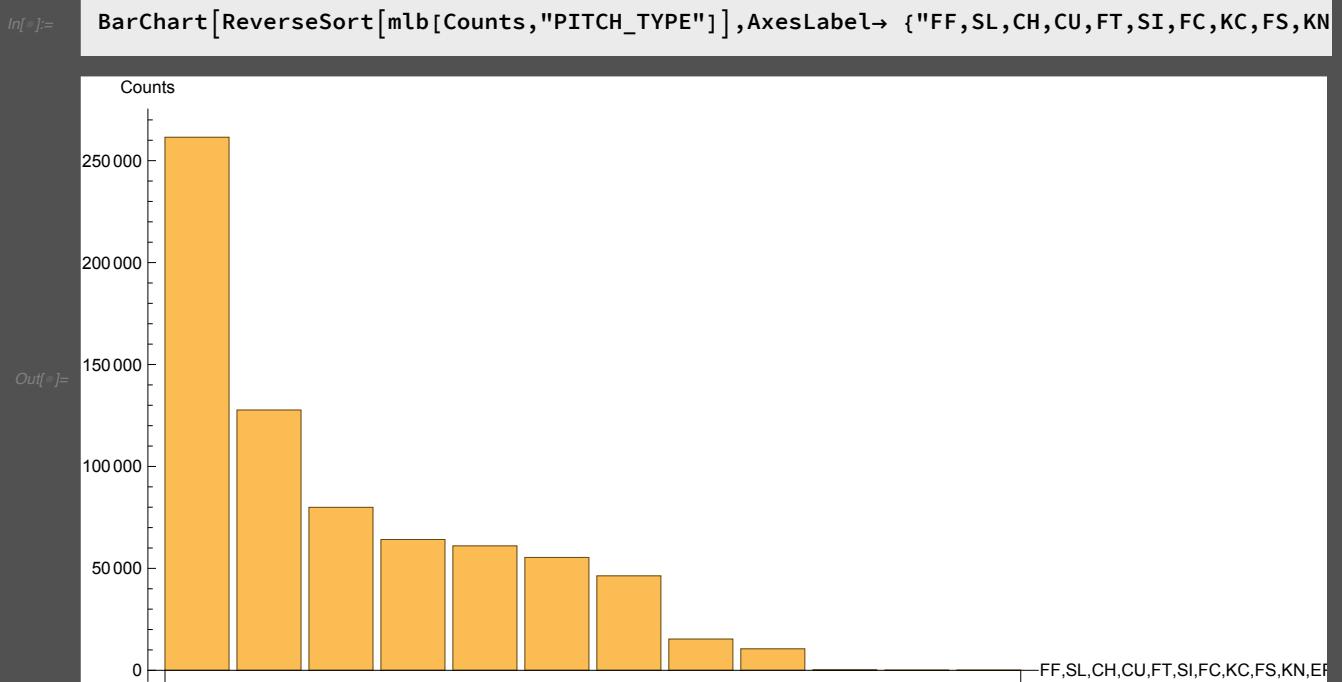
FO = *Pitchout*

In[]:=

```
ReverseSort[mlb[Counts,"PITCH_TYPE"]]
```

Out[]:=

FF	261485
SL	127680
CH	79939
CU	64166
FT	61044
SI	55361
FC	46339
KC	15324
FS	10538
KN	200
EP	49
FO	36



START SPEED (*START_SPEED*)

Velocity recorded 10 feet out of the PITCHER'S hand.

Fastest start speeds and pitch type

Slowest start speeds and pitch type

```
In[7]:= ReverseSort[mlb[Select[#"START_SPEED" <= 51 &]]]
```

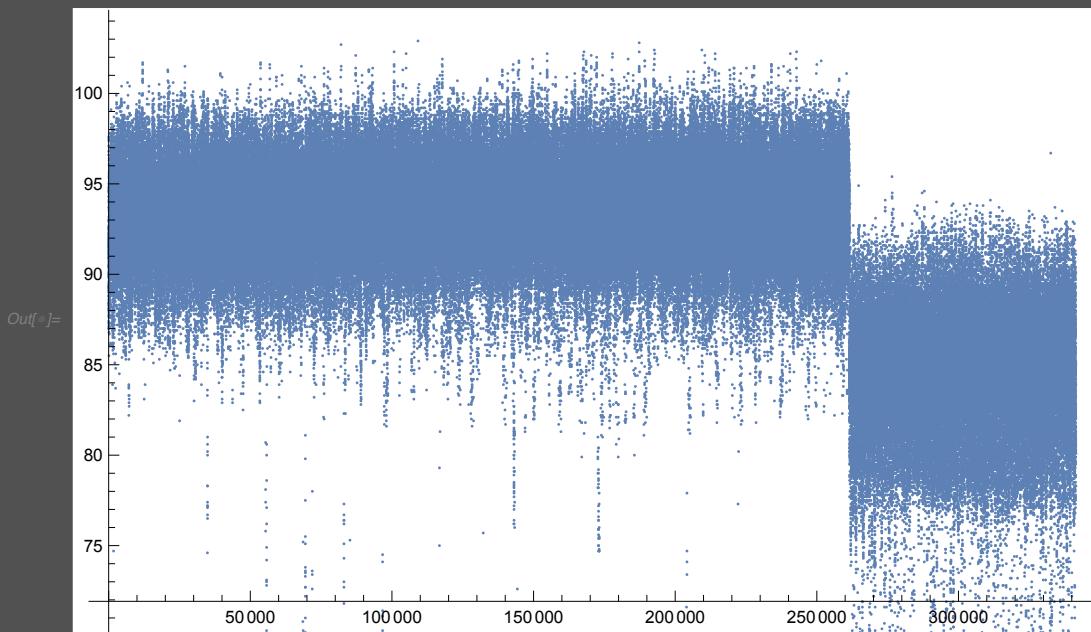
Out[7]=

START_SPEED	END_SPEED	BREAK_ANGLE	BREAK_LENGTH	PITCH_TYP	B_SCORE	B_COUNT	S_COUN
50.8	46.6	0.0	21.6	CU	19	0	0
50.7	47.8	2.4	20.4	SL	6	0	0
50.6	47.5	1.2	20.4	FF	11	2	0

◀ < columns 1–10 of 13 > ▶

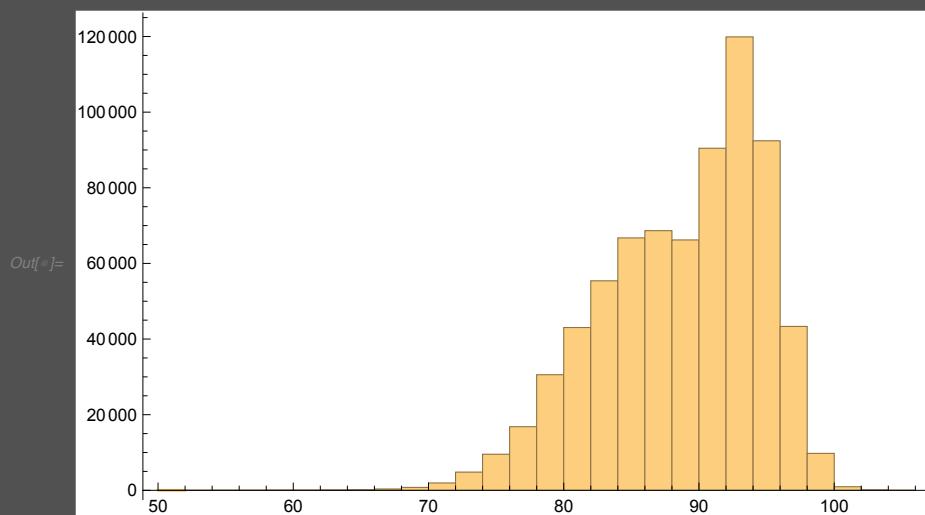
Visualization between the speeds of fastballs and changeups!

```
In[8]:= fastball=mlb[Select[#"PITCH_TYPE"=="FF" &],"START_SPEED"]//Normal;
changeup=mlb[Select[#"PITCH_TYPE"=="CH" &],"START_SPEED"]//Normal;
joinPitches=Join[fastball,changeup];
ListPlot[joinPitches]
```



Histogram of ALL Start Speeds...

```
In[¶] := mlb[Histogram,"START_SPEED"]
```



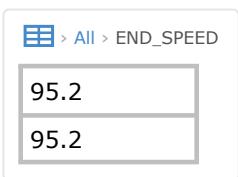
END SPEEDS (*END_SPEED*)

Velocity recorded 10 feet from the BATTER.

Four Seam Fastball

```
In[7]:= mlb[Select[#>#&#034;PITCH_TYPE">#&#034;FF"&&#034;END_SPEED">95&]]
```

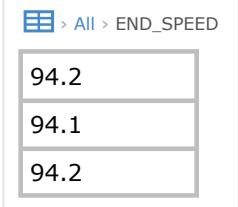
Out[7]=



Two Seam Fastball

```
In[8]:= mlb[Select[#>#&#034;PITCH_TYPE">#&#034;FT"&&#034;END_SPEED">94&]]
```

Out[8]=



Slider

```
In[9]:= mlb[Select[#>#&#034;PITCH_TYPE">#&#034;SL"&&#034;END_SPEED">90&]]
```

Out[9]=



Changeup

```
In[7]:= mlb[Select[#"PITCH_TYPE" == "CH" && #"END_SPEED" > 88.2 & ]]
```

Out[7]=

88.5

88.7

Curveball

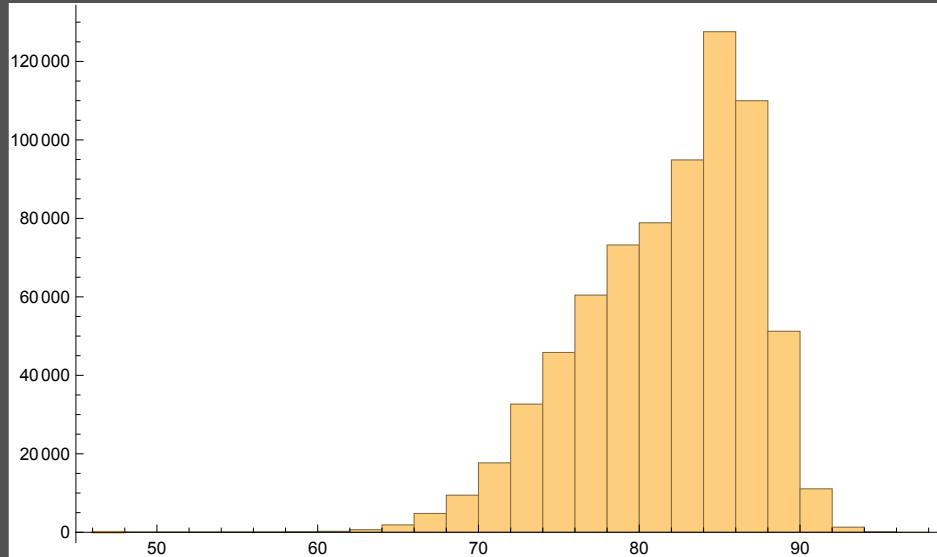
```
In[8]:= mlb[Select[#"PITCH_TYPE" == "CU" && #"END_SPEED" > 84 & ]]
```

Out[8]=

84.1

```
In[9]:= mlb[Histogram,"END_SPEED"]
```

Out[9]=



BREAK LENGTH (*BREAK_LENGTH*)

Break Length is measured from the highest point of the pitch to the lowest point of the pitch.

Curveballs are always measured closely by break length.

Mean break length of curveball: 12.2 inches

In[7]:=

```
mlb[GroupBy["PITCH_TYPE"],Mean]
```

Out[7]=

☰ > All > BREAK_LENGTH

KC	12.0164
SI	6.35622
CH	7.67622
FF	3.99299
CU	12.2092
SL	8.27212
FT	5.91402
FC	6.09825
FS	7.80068
EP	17.6327
KN	10.92
FO	9.86667

41.723 % of curveballs thrown broke more than 12 inches/1 foot.

In[8]:=

```
total=mlb[Select[#"PITCH_TYPE"=="CU"&]]//Length;
mean=mlb[Select[#"PITCH_TYPE"=="CU"&&#"BREAK_LENGTH">> 12]&]//Length;
(mean/total)*100//N
```

Out[8]= 41.723

PITCH NUMBER (PITCH_NUM)

Unlike cricket, a batter gets more than 1 try to get a hit.

Pitch number represents what number pitch they have thrown in the at-bat.

- allowed 3 strikes (batter is out)
- if pitcher throws 4 balls (batter takes their base)
- however a hitter can keep hitting the ball out of play (foul ball) and they will get more chances

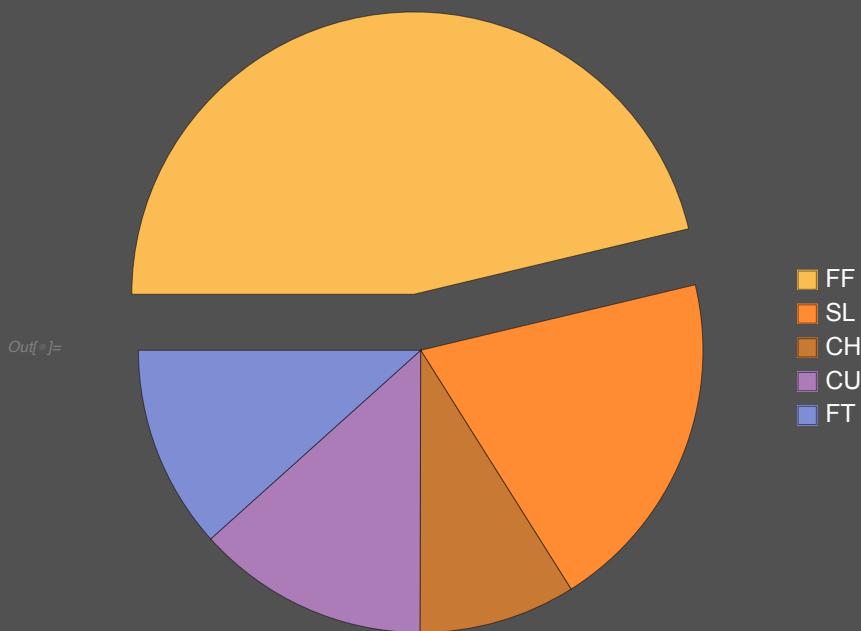
```
In[]:=
```

```
mlb[All,"PITCH_NUM"]//Normal//Max
```

```
Out[]:= 16
```

In one at-bat, a pitcher threw 16 pitches to one hitter!

```
In[4]:= firstPitch=mlb[Select[#"PITCH_NUM" == 1 &]]//Length;
firstFF = mlb[Select[#"PITCH_TYPE" == "FF" && #"PITCH_NUM" == 1 &]]//Length;
firstSL=mlb[Select[#"PITCH_TYPE" == "SL" && #"PITCH_NUM" == 1 &]]//Length;
firstCH=mlb[Select[#"PITCH_TYPE" == "CH" && #"PITCH_NUM" == 1 &]]//Length;
firstCUR= mlb[Select[#"PITCH_TYPE" == "CU" && #"PITCH_NUM" == 1 &]]//Length;
firstFT= mlb[Select[#"PITCH_TYPE" == "FT" && #"PITCH_NUM" == 1 &]]//Length;
perFF=firstFF/firstPitch*100//N;
perSL=firstSL/firstPitch*100//N;
perCH=firstCH/firstPitch*100//N;
perCUR=firstCUR/firstPitch*100//N;
perFT=firstFT/firstPitch*100//N;
piechart=PieChart[{perFF,perSL,perCH,perCUR,perFT},ChartLegends→{"FF","SL","CH","CU","FT"}]
```



Shows the percentages of each pitch of all first pitches

- 37.8% of the time, you are getting a fastball first pitch

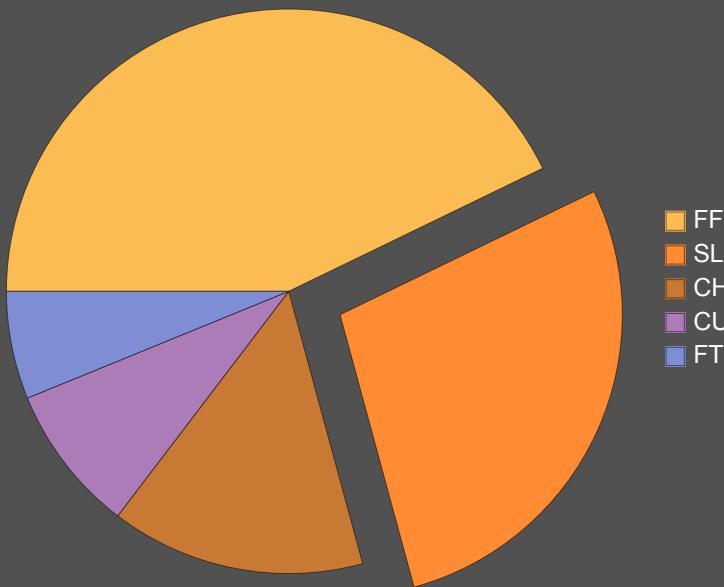
STRIKE COUNT (*S_COUNT*)

Shows how many strikes there were when the pitch is thrown.

- hitter allowed 3 strikes--> on third strike the hitter is out
- analyze pitches that are thrown with 2 strikes
- pitcher throws their best pitch to get the hitter out

```
In[1]:= strikeLength= mlb[Select[#"S_COUNT" == 2 &]] // Length;
twoStrikeFF= mlb[Select[#"PITCH_TYPE" == "FF" && #"S_COUNT" == 2&]] // Length;
twoStrikeSL=mlb[Select[#"PITCH_TYPE" == "SL" && #"S_COUNT" == 2&]] // Length;
twoStrikeCH=mlb[Select[#"PITCH_TYPE" == "CH" && #"S_COUNT" == 2&]] // Length;
twoStrikeCUR=mlb[Select[#"PITCH_TYPE" == "CU" && #"S_COUNT" == 2&]] // Length;
twoStrikeFT=mlb[Select[#"PITCH_TYPE" == "FT" && #"S_COUNT" == 2&]] // Length;
percentFF=twoStrikeFF/strikeLength*100//N;
percentSL=twoStrikeSL/strikeLength*100//N;
percentCH=twoStrikeCH/strikeLength*100//N;
percentCUR=twoStrikeCUR/strikeLength*100//N;
percentFT=twoStrikeFT/strikeLength*100//N;
twoStrikePiechart=PieChart[{percentFF,percentSL,percentCH,percentCUR,percentFT},ChartLegends→{"FF","SL","CH",
percentFT}];
```

Out[1]=



Sliders are thrown 23.9% of the time when there are two strikes!
TOUGH PITCH TO HIT

CONCLUSION

- Data Science
 - Cleaning, Manipulation, Visualization
- Storytelling
- Publish to Cloud