1. The Statcast revolution

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

# Load Aaron Judge's Statcast data
judge = pd.read_csv('datasets/judge.csv')

# Load Giancarlo Stanton's Statcast data
stanton = pd.read_csv('datasets/stanton.csv')
```

2. What can Statcast measure?

```
# Display all columns (pandas will collapse some columns if we don't
set this option)
pd.set_option('display.max_columns', None)

# Display the last five rows of the Aaron Judge file
print(judge.tail())
```

```
     pitch_type   game_date  release_speed  release_pos_x
release_pos_z  \
3431         CH  2016-08-13           85.6        -1.9659
5.9113
3432         CH  2016-08-13           87.6        -1.9318
5.9349
3433         CH  2016-08-13           87.2        -2.0285
5.8656
3434         CU  2016-08-13           79.7        -1.7108
6.1926
3435         FF  2016-08-13           93.2        -1.8476
6.0063

      player_name  batter  pitcher    events          description
spin_dir  \
3431  Aaron Judge  592450   542882       NaN                 ball
NaN
3432  Aaron Judge  592450   542882  home_run  hit_into_play_score
NaN
3433  Aaron Judge  592450   542882       NaN                 ball
NaN
3434  Aaron Judge  592450   542882       NaN                 foul
NaN
3435  Aaron Judge  592450   542882       NaN        called_strike
NaN
```

```
      spin_rate_deprecated  break_angle_deprecated
break_length_deprecated  \
3431                   NaN                     NaN
NaN
3432                   NaN                     NaN
NaN
3433                   NaN                     NaN
NaN
3434                   NaN                     NaN
NaN
3435                   NaN                     NaN
NaN

      zone                                                  des
game_type stand  \
3431  14.0                                                  NaN
R     R
3432   4.0  Aaron Judge homers (1) on a fly ball to center...
R     R
3433  14.0                                                  NaN
R     R
3434   4.0                                                  NaN
R     R
3435   8.0                                                  NaN
R     R

     p_throws home_team away_team type  hit_location   bb_type  balls
\
3431        R       NYY        TB    B            NaN       NaN      0

3432        R       NYY        TB    X            NaN  fly_ball      1

3433        R       NYY        TB    B            NaN       NaN      0

3434        R       NYY        TB    S            NaN       NaN      0

3435        R       NYY        TB    S            NaN       NaN      0


      strikes  game_year     pfx_x     pfx_z  plate_x  plate_z  on_3b
on_2b  \
3431        0       2016 -0.379108  0.370567    0.739    1.442    NaN
NaN
3432        2       2016 -0.295608  0.320400   -0.419    3.273    NaN
NaN
3433        2       2016 -0.668575  0.198567    0.561    0.960    NaN
NaN
3434        1       2016  0.397442 -0.614133   -0.803    2.742    NaN
```

```
NaN
3435        0       2016 -0.823050  1.623300   -0.273    2.471    NaN
NaN

      on_1b  outs_when_up  inning inning_topbot    hc_x   hc_y  \
3431    NaN             0       5           Bot     NaN    NaN
3432    NaN             2       2           Bot  130.45  14.58
3433    NaN             2       2           Bot     NaN    NaN
3434    NaN             2       2           Bot     NaN    NaN
3435    NaN             2       2           Bot     NaN    NaN

      tfs_deprecated  tfs_zulu_deprecated  pos2_person_id  umpire  \
3431             NaN                  NaN        571912.0     NaN
3432             NaN                  NaN        571912.0     NaN
3433             NaN                  NaN        571912.0     NaN
3434             NaN                  NaN        571912.0     NaN
3435             NaN                  NaN        571912.0     NaN

               sv_id    vx0      vy0    vz0     ax      ay      az
sz_top  \
3431  160813_144259  6.960 -124.371 -4.756 -2.821  23.634 -30.220
3.93
3432  160813_135833  4.287 -127.452 -0.882 -1.972  24.694 -30.705
4.01
3433  160813_135815  7.491 -126.665 -5.862 -6.393  21.952 -32.121
4.01
3434  160813_135752  1.254 -116.062  0.439  5.184  21.328 -39.866
4.01
3435  160813_135736  5.994 -135.497 -6.736 -9.360  26.782 -13.446
4.01

      sz_bot  hit_distance_sc  launch_speed  launch_angle
effective_speed  \
3431    1.82              NaN           NaN           NaN
84.459
3432    1.82            446.0         108.8        27.410
86.412
3433    1.82              NaN           NaN           NaN
86.368
3434    1.82              9.0          55.8       -24.973
77.723
3435    1.82              NaN           NaN           NaN
92.696

      release_spin_rate  release_extension  game_pk  pos1_person_id  \
3431             1552.0              5.683   448611        542882.0
3432             1947.0              5.691   448611        542882.0
3433             1761.0              5.721   448611        542882.0
3434             2640.0              5.022   448611        542882.0
3435             2271.0              6.068   448611        542882.0
```

```
      pos2_person_id.1  pos3_person_id  pos4_person_id  pos5_person_id
\
3431          571912.0        543543.0        523253.0        446334.0

3432          571912.0        543543.0        523253.0        446334.0

3433          571912.0        543543.0        523253.0        446334.0

3434          571912.0        543543.0        523253.0        446334.0

3435          571912.0        543543.0        523253.0        446334.0


      pos6_person_id  pos7_person_id  pos8_person_id
pos9_person_id  \
3431        622110.0        545338.0        595281.0        543484.0

3432        622110.0        545338.0        595281.0        543484.0

3433        622110.0        545338.0        595281.0        543484.0

3434        622110.0        545338.0        595281.0        543484.0

3435        622110.0        545338.0        595281.0        543484.0


      release_pos_y  estimated_ba_using_speedangle  \
3431        54.8144                           0.00
3432        54.8064                           0.98
3433        54.7770                           0.00
3434        55.4756                           0.00
3435        54.4299                           0.00

      estimated_woba_using_speedangle  woba_value  woba_denom
babip_value  \
3431                            0.000         NaN         NaN
NaN
3432                            1.937         2.0         1.0
0.0
3433                            0.000         NaN         NaN
NaN
3434                            0.000         NaN         NaN
NaN
3435                            0.000         NaN         NaN
NaN

      iso_value  launch_speed_angle  at_bat_number  pitch_number
3431        NaN                 NaN             36             1
```

```
3432        3.0                 6.0             14              4
3433        NaN                 NaN             14              3
3434        NaN                 1.0             14              2
3435        NaN                 NaN             14              1
```

3. Aaron Judge and Giancarlo Stanton, prolific sluggers

```
# All of Aaron Judge's batted ball events in 2017
judge_events_2017 =judge.loc[judge['game_year']==2017].events
print("Aaron Judge batted ball event totals, 2017:")
print(judge_events_2017.value_counts())

# All of Giancarlo Stanton's batted ball events in 2017
stanton_events_2017 = stanton.loc[stanton['game_year']==2017].events
print("\nGiancarlo Stanton batted ball event totals, 2017:")
print(stanton_events_2017.value_counts())
```

```
Aaron Judge batted ball event totals, 2017:
strikeout                   207
field_out                   146
walk                        116
single                       75
home_run                     52
double                       24
grounded_into_double_play    15
intent_walk                  11
force_out                    11
hit_by_pitch                  5
fielders_choice_out           4
field_error                   4
sac_fly                       4
triple                        3
strikeout_double_play         1
Name: events, dtype: int64

Giancarlo Stanton batted ball event totals, 2017:
field_out                   239
strikeout                   161
single                       77
walk                         72
home_run                     59
double                       32
intent_walk                  13
grounded_into_double_play    13
hit_by_pitch                  7
force_out                     7
field_error                   5
sac_fly                       3
fielders_choice_out           2
```

```
strikeout_double_play          2
pickoff_1b                     1
Name: events, dtype: int64
```

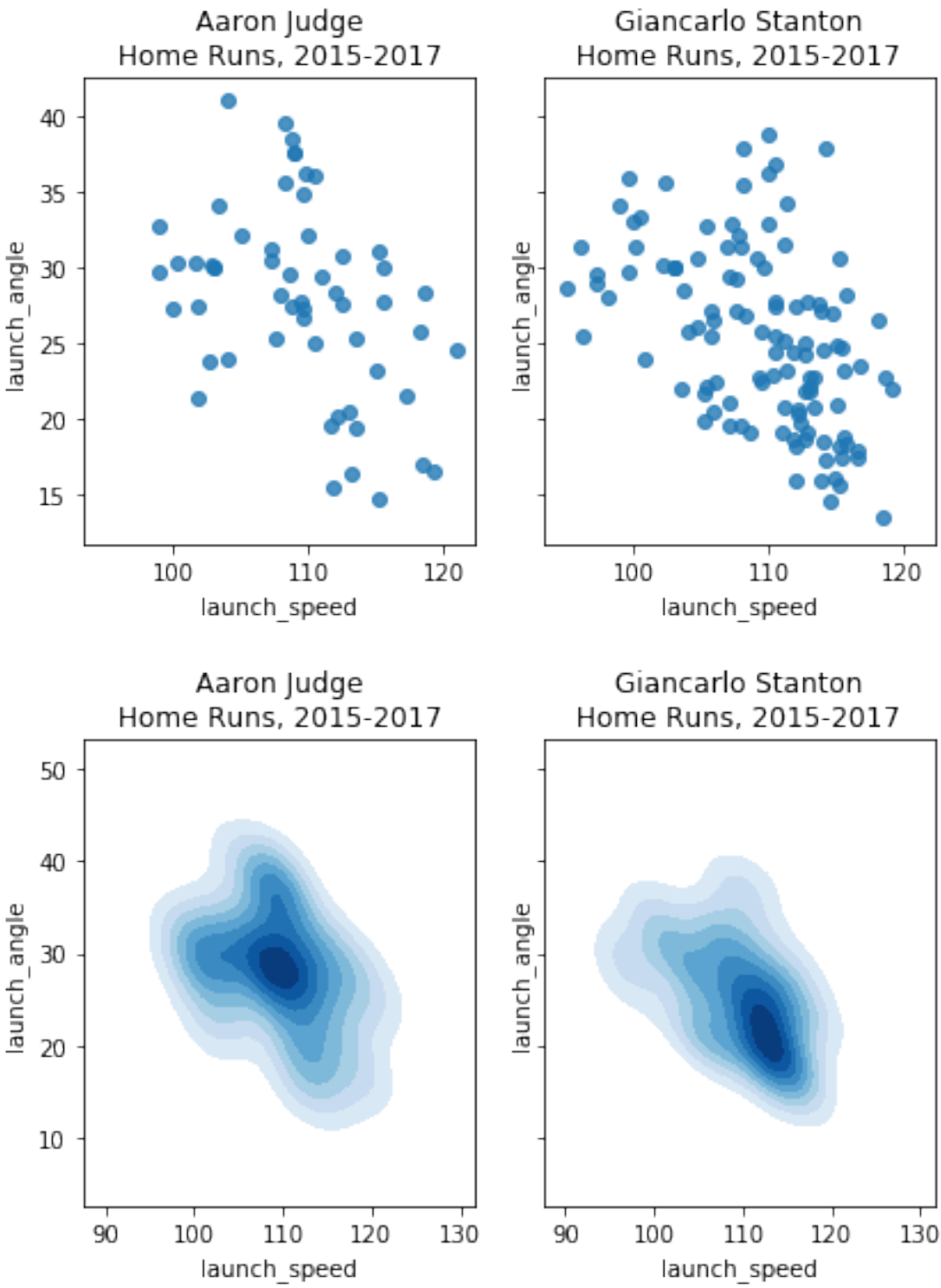
4. Analyzing home runs with Statcast data

```
# Filter to include home runs only
judge_hr = judge.loc[judge['events']=='home_run']
stanton_hr = stanton.loc[stanton['events']=='home_run']

# Create a figure with two scatter plots of launch speed vs. launch
angle, one for each player's home runs
fig1, axs1 = plt.subplots(ncols=2, sharex=True, sharey=True)
sns.regplot(x=judge_hr['launch_speed'], y=judge_hr['launch_angle'],
fit_reg=False, color='tab:blue', data=judge_hr,
ax=axs1[0]).set_title('Aaron Judge\nHome Runs, 2015-2017')
sns.regplot(x=stanton_hr['launch_speed'],
y=stanton_hr['launch_angle'], fit_reg=False, color='tab:blue',
data=stanton_hr, ax=axs1[1]).set_title('Giancarlo Stanton\nHome Runs,
2015-2017')

# Create a figure with two KDE plots of launch speed vs. launch angle,
one for each player's home runs
fig2, axs2 = plt.subplots(ncols=2, sharex=True, sharey=True)
sns.kdeplot(judge_hr['launch_speed'], judge_hr['launch_angle'],
cmap="Blues", shade=True, shade_lowest=False,
ax=axs2[0]).set_title('Aaron Judge\nHome Runs, 2015-2017')
sns.kdeplot(stanton_hr['launch_speed'], stanton_hr['launch_angle'],
cmap="Blues", shade=True, shade_lowest=False,
ax=axs2[1]).set_title('Giancarlo Stanton\nHome Runs, 2015-2017')
```

```
Text(0.5,1,'Giancarlo Stanton\nHome Runs, 2015-2017')
```

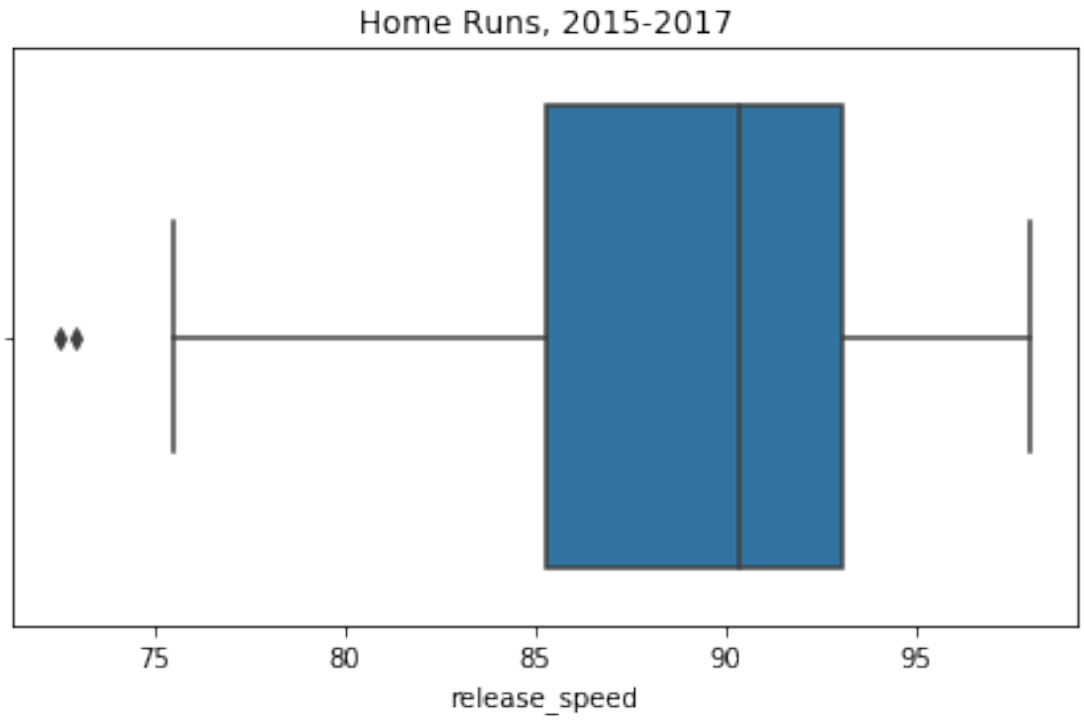


5. Home runs by pitch velocity

```
# Combine the Judge and Stanton home run DataFrames for easy boxplot
plotting
judge_stanton_hr = pd.concat([judge_hr, stanton_hr])
#print(type(judge_stanton_hr['release_speed']))
```

```
# Create a boxplot that describes the pitch velocity of each player's
home runs
sns.boxplot(x=judge_stanton_hr['release_speed'],
color='tab:blue').set_title('Home Runs, 2015-2017')
```

```
Text(0.5,1,'Home Runs, 2015-2017')
```



6. Home runs by pitch location (I)

```
def assign_x_coord(row):
    """
    Assigns an x-coordinate to Statcast's strike zone numbers. Zones
11, 12, 13,
    and 14 are ignored for plotting simplicity.
    """
    # Left third of strike zone
    if row.zone in [1, 4, 7]:
        return 1
    # Middle third of strike zone
    if row.zone in [2, 5, 8]:
        return 2
    # Right third of strike zone
    if row.zone in [3, 6, 9]:
        return 3
```

7. Home runs by pitch location (II)

```
def assign_y_coord(row):
    """
```

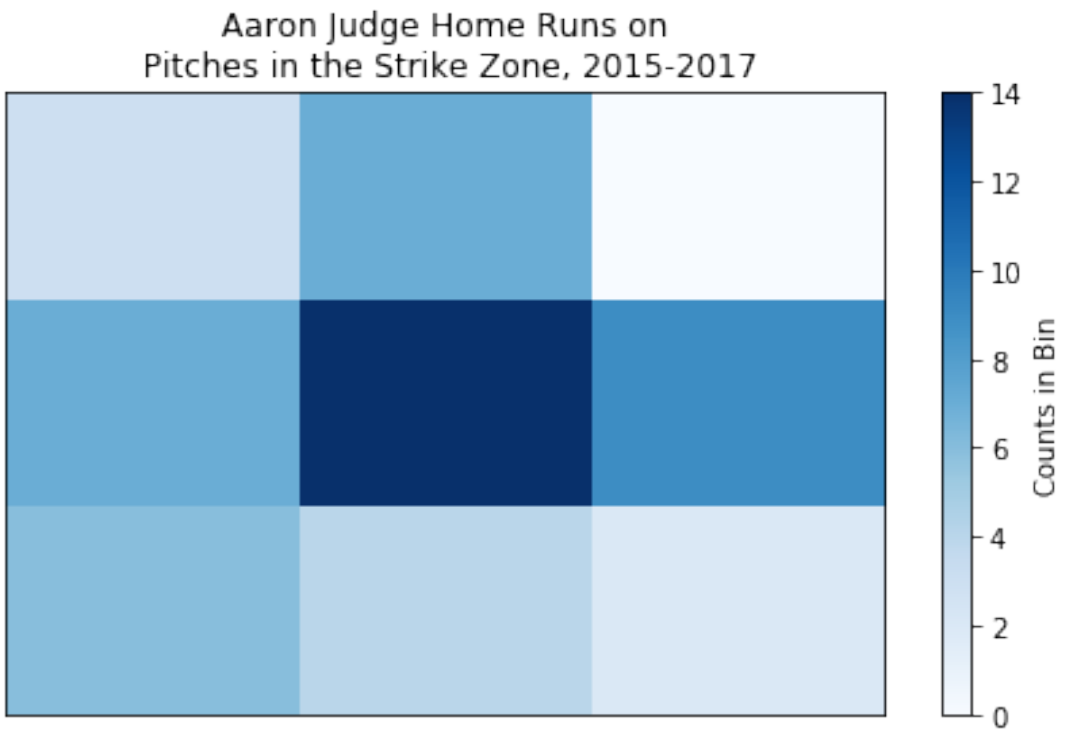
```
    Assigns a y-coordinate to Statcast's strike zone numbers. Zones
11, 12, 13,
    and 14 are ignored for plotting simplicity.
    """
    # Upper third of strike zone
    if row.zone in [1, 2, 3]:
        return 3
    # Middle third of strike zone
    if row.zone in [4, 5, 6]:
        return 2
    # Lower third of strike zone
    if row.zone in [7, 8, 9]:
        return 1
```

8. Aaron Judge's home run zone

```
# Zones 11, 12, 13, and 14 are to be ignored for plotting simplicity
judge_strike_hr = judge_hr.copy().loc[judge_hr.zone <= 9]

# Assign Cartesian coordinates to pitches in the strike zone for Judge
home runs
judge_strike_hr['zone_x'] = judge_strike_hr.apply(assign_x_coord,
axis=1)
judge_strike_hr['zone_y'] = judge_strike_hr.apply(assign_y_coord,
axis=1)

# Plot Judge's home run zone as a 2D histogram with a colorbar
plt.hist2d(judge_strike_hr['zone_x'], judge_strike_hr['zone_y'], bins
= 3, cmap='Blues')
plt.title('Aaron Judge Home Runs on\n Pitches in the Strike Zone,
2015-2017')
plt.gca().get_xaxis().set_visible(False)
plt.gca().get_yaxis().set_visible(False)
cb = plt.colorbar()
cb.set_label('Counts in Bin')
```

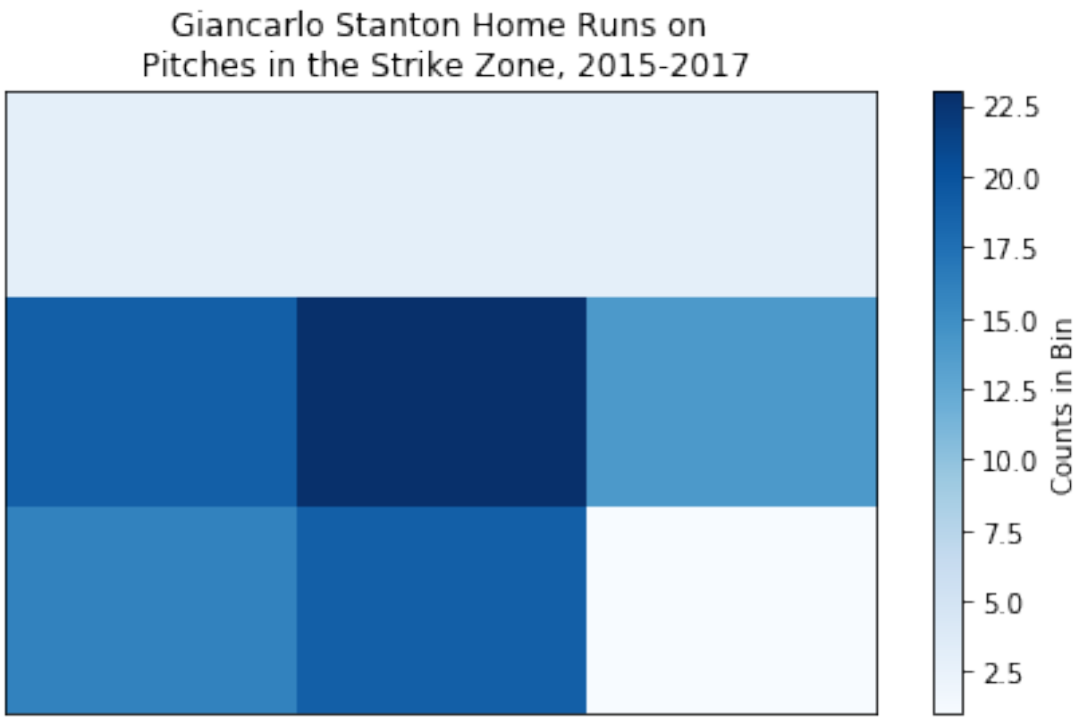


9. Giancarlo Stanton's home run zone

```
# Zones 11, 12, 13, and 14 are to be ignored for plotting simplicity
stanton_strike_hr = stanton_hr.copy().loc[stanton_hr.zone <= 9]

# Assign Cartesian coordinates to pitches in the strike zone for
Stanton home runs
stanton_strike_hr['zone_x']=stanton_strike_hr.apply(assign_x_coord,
axis=1)
stanton_strike_hr['zone_y']=stanton_strike_hr.apply(assign_y_coord,
axis=1)

# Plot Stanton's home run zone as a 2D histogram with a colorbar
plt.hist2d(stanton_strike_hr['zone_x'], stanton_strike_hr['zone_y'],
bins=3, cmap='Blues' )
plt.title('Giancarlo Stanton Home Runs on\n Pitches in the Strike
Zone, 2015-2017')
plt.gca().get_xaxis().set_visible(False)
plt.gca().get_yaxis().set_visible(False)
cb = plt.colorbar()
cb.set_label('Counts in Bin')
```



10. Should opposing pitchers be scared?

```
# Should opposing pitchers be wary of Aaron Judge and Giancarlo
Stanton
should_pitchers_be_scared = True
```