

Major League Baseball Player-Skill Analysis

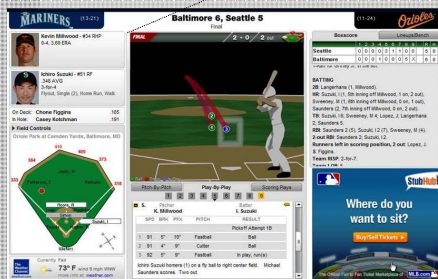
with Pitch F/x Data and ArcGIS

By: Grant Humphries

Gameday and Pitch F/x

Pitch F/x (a GIS)

Gameday Application



Attribute Table

FINAL 2 - 0 2 out

Ball Strike In Play

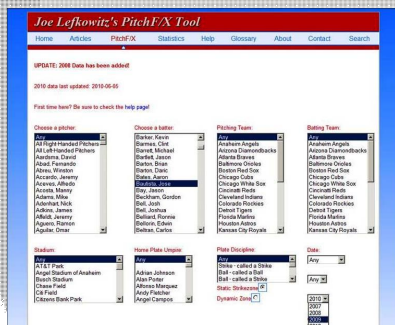
Pitch-By-Pitch Play-By-Play Scoring Plays

Pitcher K. Millwood				Batter I. Suzuki	
SPD	BRK	PFx	PITCH	RESULT	
Pickoff Attempt 1B					
1	91	5"	10"	Fastball	Ball
2	91	4"	9"	Cutter	Ball
3	92	5"	9"	Fastball	In play, run(s)

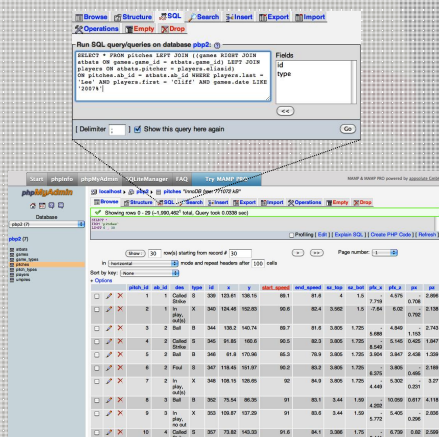
Ichiro Suzuki homers (1) on a fly ball to right center field. Michael Saunders scores. Two out.

Acquiring and Organizing the Data

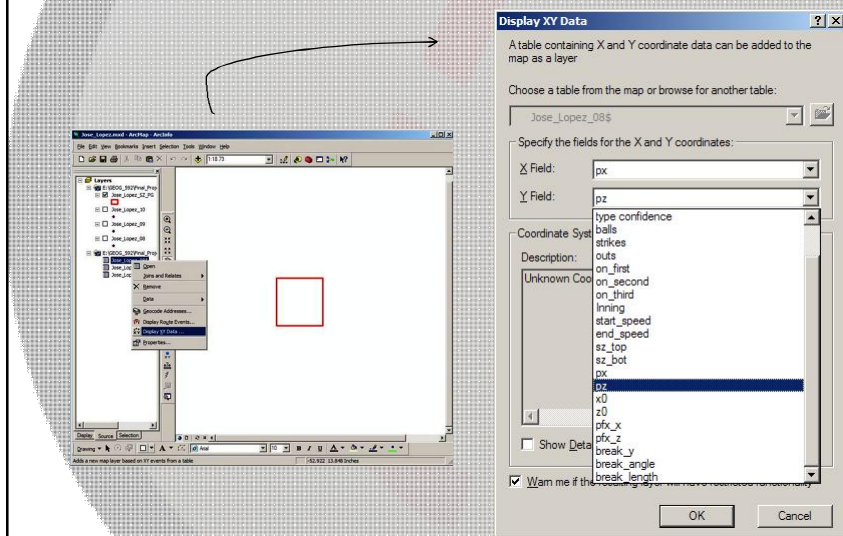
Joe Lefkowitz's Free Online Pitch F/x Tool



Queries in MySQL Database



Importing Data into ArcMap



Research Questions

- What is the root or underlying cause when a player has a large change in their production?
 - Is the change tied to luck or are they exhibiting new skills?
- What is the player's true talent level?
- What level of performance can be expected from a player going forward?

Case 1:

Adrian Gonzalez's

Improved Walk Rate in 2009

Plate Discipline Statistics

<u>Season</u>	<u>Team</u>	<u>PA</u>	<u>BB</u>	<u>BB%</u>
2008	Padres	700	74	10.60%
2009	Padres	681	119	17.50%

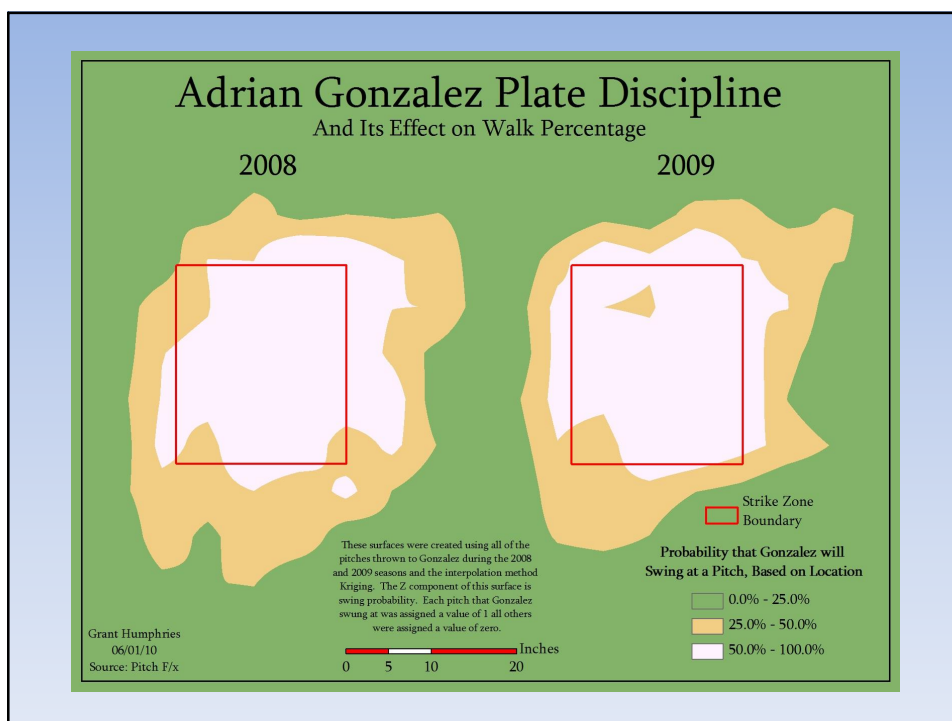
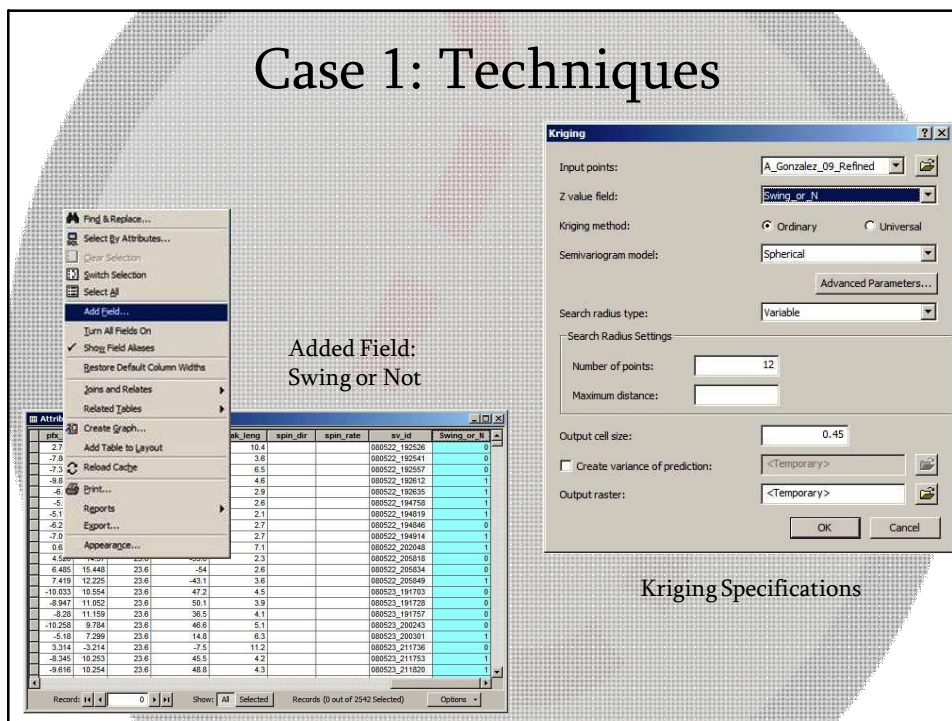
General Offense Statistics

<u>2B</u>	<u>3B</u>	<u>HR</u>	<u>OBP</u>	<u>SLG</u>	<u>ISO</u>	<u>OPS</u>	<u>wOBA</u>	<u>WAR</u>
32	1	36	0.361	0.510	0.231	0.871	0.368	3.9
27	2	40	0.407	0.551	0.274	0.958	0.402	6.6

Stats via



Case 1: Techniques



Case 1: Conclusions

- Has Approach Changed
Yes
- New Skills Displayed?
Yes
- Sustainable?
Yes

Case 2:

Joel Piniero's

Increase in Groundballs in the 2009 Season

Batted Ball Statistics

<u>Season</u>	<u>Team</u>	<u>IP</u>	<u>GB</u>	<u>LD%</u>	<u>FB%</u>	<u>GB%</u>	<u>GB/FB</u>	<u>HR/9</u>	<u>HR/FB</u>
2008	Cardinals	148.2	250	21.7%	29.7%	48.5%	1.63	1.33	14.4%
2009	Cardinals	214.0	427	15.7%	23.8%	60.5%	2.54	0.46	6.5%

General Pitching Statistics

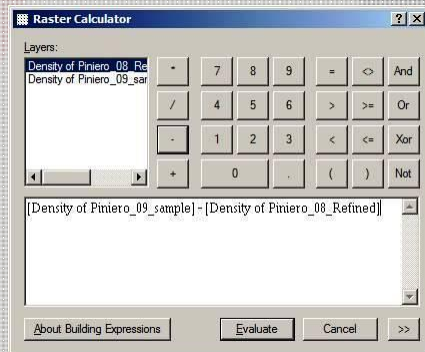
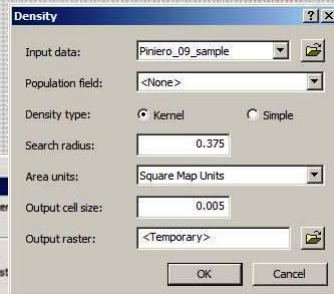
<u>K/9</u>	<u>BB/9</u>	<u>ERA</u>	<u>FIP</u>	<u>xFIP</u>	<u>WAR</u>
4.9	2.1	5.15	4.71	4.19	0.9
4.4	1.1	3.49	3.27	3.68	4.8

Stats via



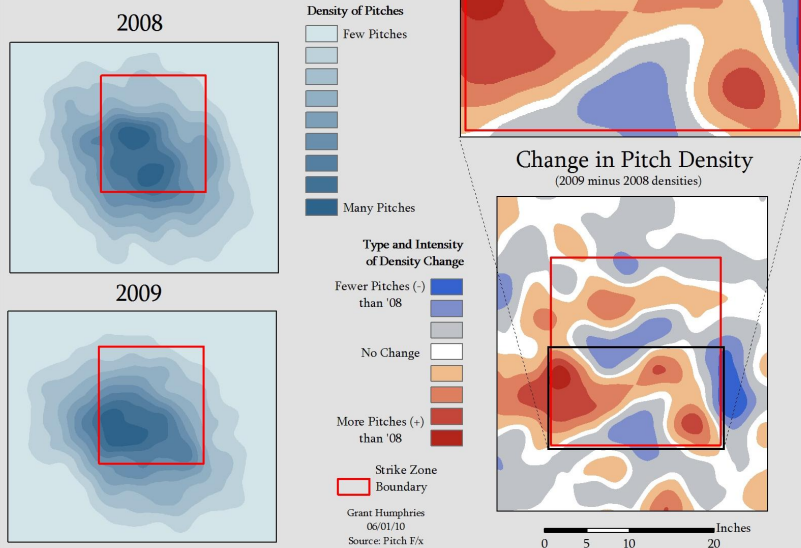
Case 2: Techniques

Created Density Maps
of Pitch Locations



Used Raster Calculator to
Determine Change in
Density from '08 to '09

Piniero's Pitch Location And Its Effect on Groundball Percentage



Case 2: Conclusions

- Has Approach Changed?

Yes

- New Skills Displayed?

???

- Sustainable?

Yes

Case 3:

John Danks Collapse

In the Second Half of the 2009 Season

General Pitching Statistics

<u>Season</u>	<u>Team</u>	<u>K/9</u>	<u>BB/9</u>	<u>HR/9</u>	<u>FIP</u>	<u>xFIP</u>	<u>GB%</u>	<u>FB%</u>
2009 1/2	White Sox	8.3	3.32	1.14	4.11	3.94	46.4%	36.9%
2009 2/2	White Sox	5.6	3.3	1.38	5.06	4.95	42.6%	43.8%

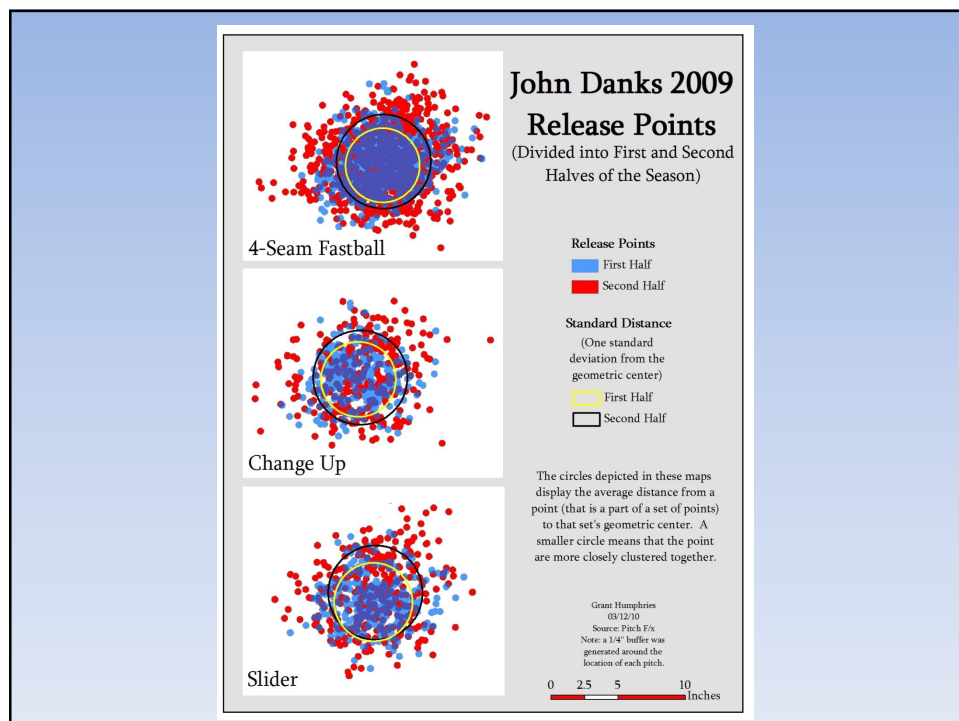
Advanced Pitch Statistics

<u>1st Half</u>				
	<u>Swinging Strike %</u>	<u>Average Velo</u>	<u>Average pfx</u>	<u>Average brk</u>
Change Up	22.8%	82.9 mph	11.95"	-
4-Seam FB	8.2%	91.0 mph	-	-
Slider	17.5%	86.1 mph	-	6.67"
<u>2nd Half</u>				
	<u>Swinging Strike %</u>	<u>Average Velo</u>	<u>Average pfx</u>	<u>Average brk</u>
Change Up	17.3%	82.1 mph	11.55"	-
4-Seam FB	3.1%	89.9 mph	-	-
Slider	10.5%	85.2 mph	-	6.52"

Case 3: Techniques

Created a 1/4" Buffer around Pitch Release Points

Used Standard Distance Tool to Determine 'Clusteredness' of Release Points



Case 3: Conclusions

- Has Approach Changed?

Yes

- New Skills Displayed?

No

- Sustainable?

No

Case 4:

Jose Bautista's Breakout

Offensive Season in 2010

Advanced Offensive Statistics

Season	Team	OBP	SLG	ISO	OPS	wOBA	wRC+	WAR
2009	Blue Jays	0.349	0.408	0.173	0.757	0.339	107	1.9
2010	Blue Jays	0.377	0.605	0.355	0.982	0.415	161	2.0

Basic Offensive Statistics

PA	BB	2B	3B	HR	BB%
404	56	13	3	13	13.9%
244	39	13	2	18	16.0%

Stats via



Case 4: Techniques

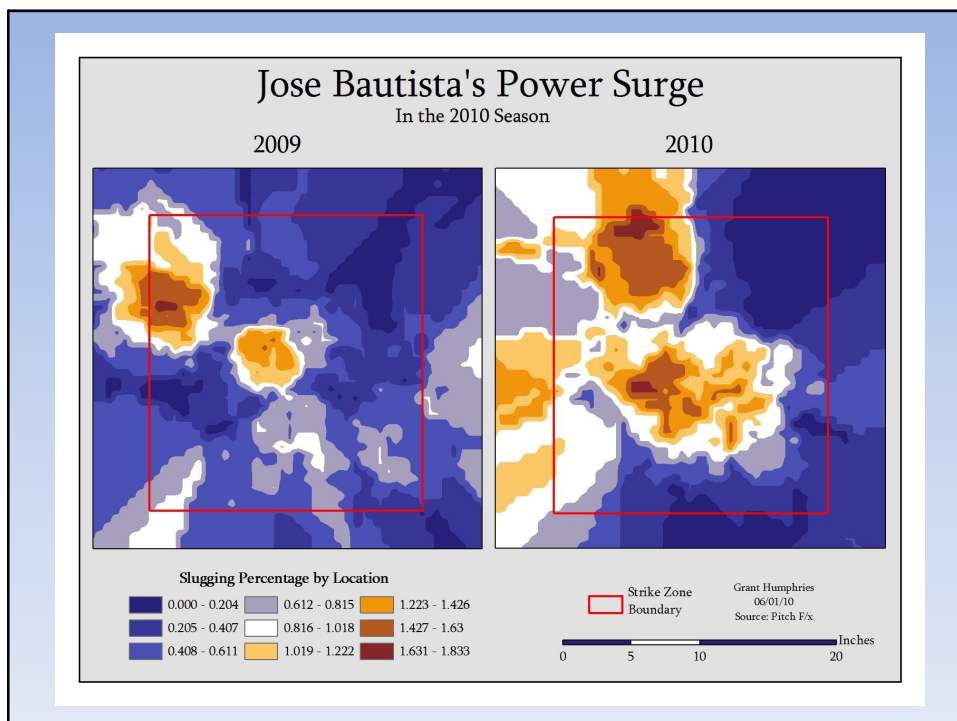
Created a New Field and Assigned Figures to Each Pitch Based on Slugging Value (Power)

The screenshot displays the ArcGIS interface with several windows open:

- Attributes of Bautista_09_BIP**: A table showing data for various pitches. The columns are `sv_id`, `SZ_x`, `SZ_y`, and `Slug_Value`. The data is as follows:

sv_id	SZ_x	SZ_y	Slug_Value
090909_195341	0	0	0
090909_210563	0	0	0
090910_131015	0	0	2
090910_134246	0	0	4
090911_200303	0	0	0
090913_134214	0	0	1
090913_143022	0	0	0
090913_150224	0	0	0
090914_194334	0	0	0
090914_204003	0	0	0
090914_213237	0	0	0
090915_202832	0	0	0
090915_224000	0	0	4
090916_195326	0	0	4
090916_202243	0	0	0
090918_193946	0	0	1
090918_212822	0	0	1
090918_224550	0	0	0
090919_202230	0	0	0
090919_205647	0	0	0
- Field Calculator**: A window for calculating new fields. The `Slug_Value` field is selected.
- Kriging**: A window for interpolating a surface. The `Slug_Value` field is selected as the input. The `Kriging method` is set to `Ordinary`, and the `Semivariogram model` is set to `Spherical`.

Interpolated a Surface with Slugging as the Z-Component



Case 4: Conclusions

- Has Approach Changed?
???
- New Skills Displayed?
Yes
- Sustainable?
??? - Need More Data

Questions?

Contact Info:
grant.humphries@gmail.com