

InjuryFX: Measuring Injury Impact on Baseball Performance



Joe Izenman



Shawn Kessler



Giles Ochs

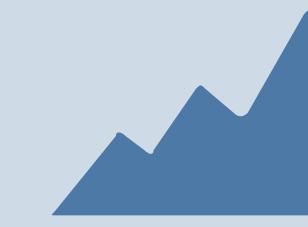
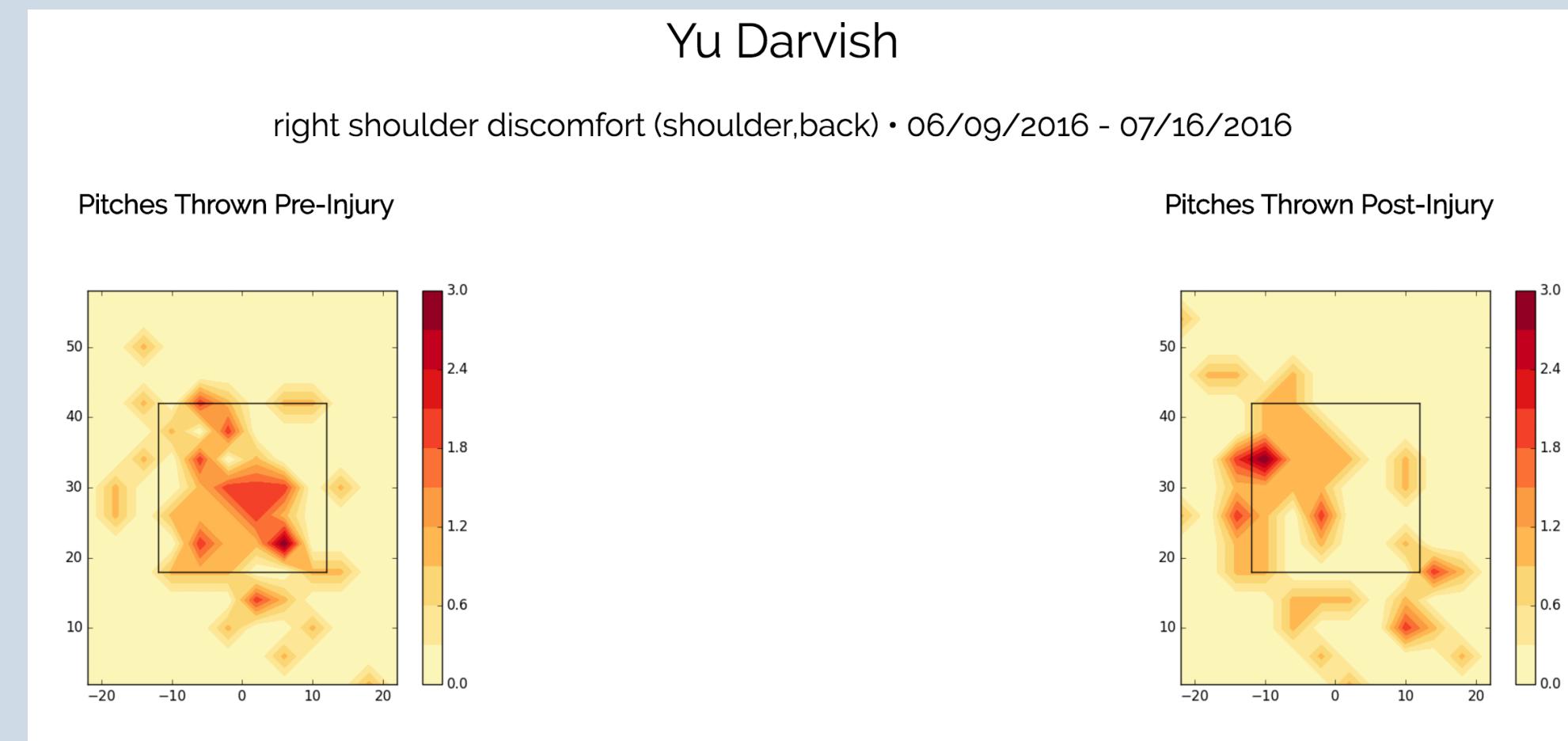


Jessica Sanders

W 205 FINAL PRESENTATION | DECEMBER 5, 2016

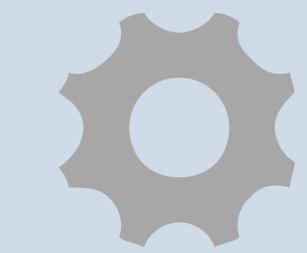


InjuryFX Overview



GOAL

Develop a tool that can be used to analyze, and project, the impact of a injury on batters and pitchers in the MLB



CHALLENGE

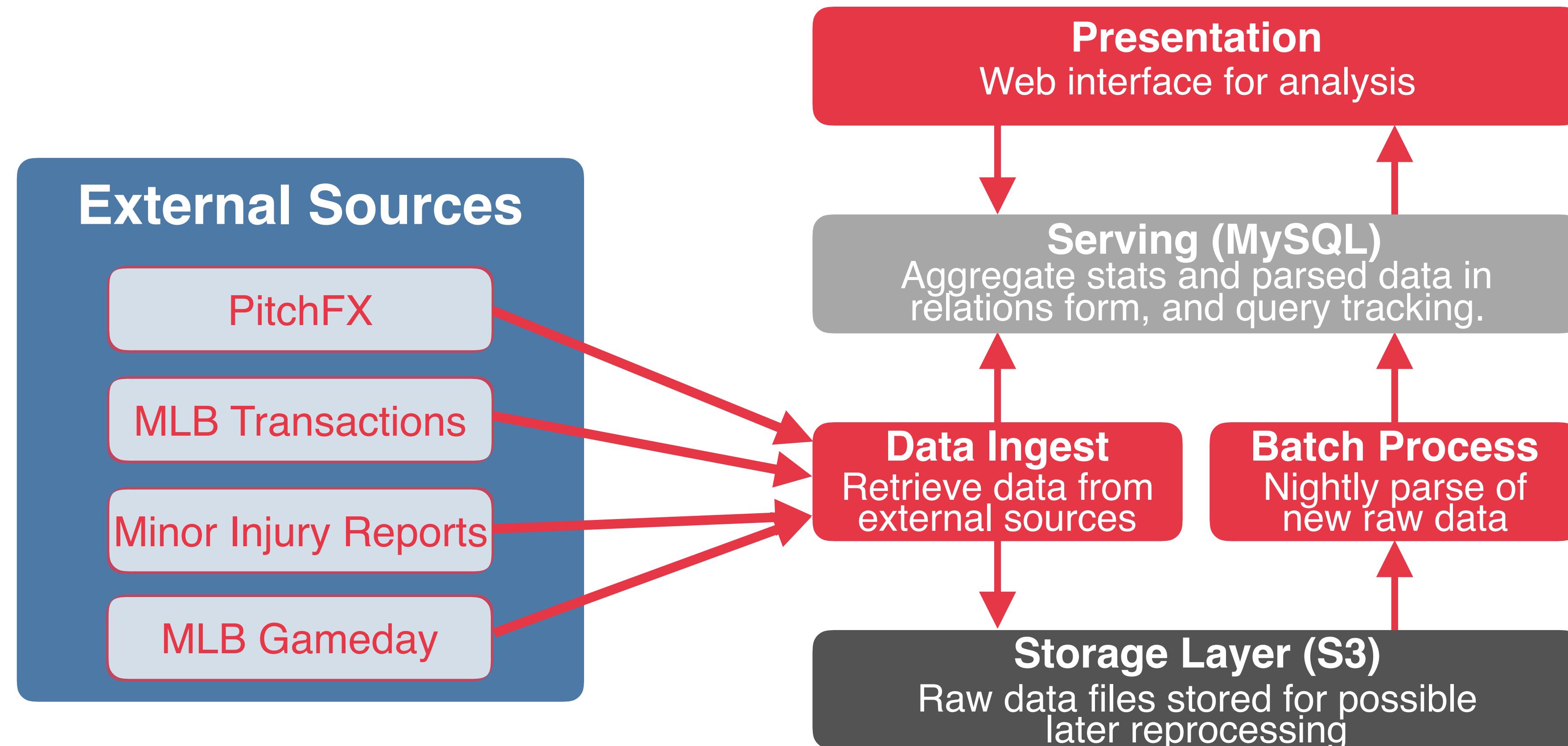
Combining disparate data sources to form a central repository of pitching and batting data.



TOOL

Build side-by-side heat maps of players before and after injury, easy to consume aggregate stats, and time series graphs that display how pitching and hitting change over time.

Architecture & Dependencies



Dependencies

- mysqlDb
- Apache2
- python2.7
- flask
- numpy
- pandas
- Sqlalchemy
- pymysql
- BeautifulSoup

Data Sources: PitchF/X

Index of /components/game/mlb/year_2014/month_10/day_29/gid_2014_10_29_sfmlb_kcamlb_1

- [Parent Directory](#)
- [atv_game_events.xml](#)
- [atv_preview.xml](#)
- [atv_preview_noscores.xml](#)
- [atv_runScoringPlays.xml](#)
- [batters/](#)
- [bench.xml](#)
- [benchO.xml](#)
- [bis_boxscore.xml](#)
- [boxscore.json](#)
- [boxscore.xml](#)
- [emailSource.xml](#)
- [eventLog.xml](#)
- [feed_selection.plist](#)
- [game.xml](#)
- [game_events.json](#)
- [game_events.plist](#)
- [game_events.xml](#)
- [game_feed.RECENT.json](#)
- [game_feed.json](#)
- [gamecenter.xml](#)
- [gameday_Syn.xml](#)
- [inning/](#)
- [linescore.json](#)
- [linescore.xml](#)
- [media/](#)
- [miniscoreboard.xml](#)
- [notifications/](#)
- [pitchers/](#)
- [players.xml](#)
- [plays.json](#)
- [plays.xml](#)
- [pregumbo.json](#)
- [premium/](#)
- [preview.plist](#)
- [rawboxscore.xml](#)
- [runScoringPlays.plist](#)

Index of /components/game/mlb/year_2014/month_10/day_29/gid_2014_10_29_sfmlb_kcamlb_1/batters

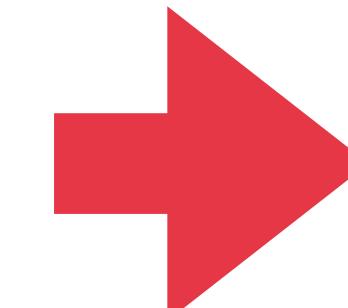
- [Parent Directory](#)
- [218596.xml](#)
- [285064.xml](#)
- [340192.xml](#)
- [346793.xml](#)
- [408241.xml](#)
- [408299.xml](#)
- [425386.xml](#)
- [425545.xml](#)
- [425657.xml](#)
- [430630.xml](#)
- [430912.xml](#)
- [433586.xml](#)
- [433589.xml](#)
- [434604.xml](#)
- [434624.xml](#)
- [435078.xml](#)
- [444876.xml](#)
- [448170.xml](#)
- [448306.xml](#)
- [450306.xml](#)
- [451584.xml](#)

This XML file does not appear to have any style information associated with it. The document tree is shown below.

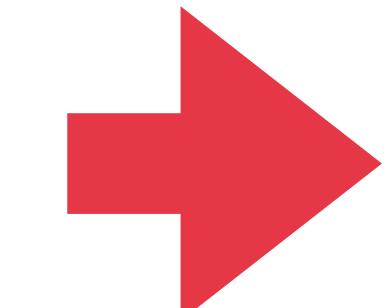
```
<!--
Copyright 2016 MLB Advanced Media, L.P. Use of any content on this page acknowledges agreement to the terms posted here http://gdx.mlb.com/components/copyright.txt
-->
<Player team="sf" id="433589" pos="P" type="batter" first_name="Yusmeiro" last_name="Petit" jersey_number="52" height="6-1" weight="255" bats="R" throws="R" dob="11/22/1984">
<season avg=".040" ab="25" h="1" bb="1" so="13" r="2" sb="0" cs="0" hr="0" rbi="0" ops=".117"/>
<series des="World Series" avg="1.000" ab="1" h="1" bb="0" so="0" r="0" sb="0" cs="0" hr="0" rbi="0" ops="2.000"/>
<month des="Sept/Oct" avg=".000" ab="0" h="0" bb="0" so="0" r="0" sb="0" cs="0" hr="0" rbi="0" ops=".000"/>
<Team des="Away" avg=".000" ab="12" h="0" bb="0" so="6" r="0" sb="0" cs="0" hr="0" rbi="0" ops=".000"/>
<Empty avg=".053" ab="19" h="1" bb="1" so="10" r="0" sb="0" cs="0" hr="0" rbi="0" ops=".153"/>
<Men_On avg=".000" ab="6" h="0" bb="0" so="3" r="2" sb="0" cs="0" hr="0" rbi="0" ops=".000"/>
<RISP avg=".000" ab="4" h="0" bb="0" so="1" r="2" sb="0" cs="0" hr="0" rbi="0" ops=".000"/>
<Loaded avg=".000" ab="0" h="0" bb="0" so="0" r="1" sb="0" cs="0" hr="0" rbi="0" ops=".000"/>
<vs_LHP avg=".167" ab="6" h="1" bb="1" so="5" r="2" sb="0" cs="0" hr="0" rbi="0" ops=".453"/>
<vs_RHP avg=".000" ab="19" h="0" bb="0" so="8" r="0" sb="0" cs="0" hr="0" rbi="0" ops=".000"/>
<vs_P des="Holland, G" avg=".000" ab="0" h="0" bb="0" so="0" r="0" sb="0" cs="0" hr="0" rbi="0" ops=".000"/>
<vs_P5 des="Holland, G" avg=".000" ab="0" h="0" bb="0" so="0" r="0" sb="0" cs="0" hr="0" rbi="0" ops=".000"/>
<Pitch loves="--" hates="--"/>
<faced/>
<atbats/>
</Player>
```

Ingestion & Processing

PitchF/X



PyGameDay
(open source library)



MySQL DB
(5 tables: atbat,
hitchart, game, pitch,
and player)

Data Sources: MLB Transactions

Ingestion & Processing

JSON API

Date	Transaction
08/31/16	New York Yankees traded RF Ben Gamel to Seattle Mariners for RHP Jio Orozco and RHP Juan De Paula. Arizona Diamondbacks optioned RHP Vicente Campos to Reno Aces.
	Seattle Mariners signed free agent RHP Luis Alcantara to a minor league contract.
	Los Angeles Dodgers signed free agent RHP Luis Cabrera to a minor league contract. Chicago White Sox signed free agent RHP Ruddy Rodriguez to a minor league contract.
	Miami Marlins placed RHP David Phelps on the 15-day disabled list retroactive to August 27, 2016. Left oblique strain.
	Miami Marlins signed free agent C Stephen McGee to a minor league contract. Toronto Blue Jays signed free agent OF Quintin Berry to a minor league contract.
	Toronto Blue Jays signed free agent C Josh Thole.
	Cincinnati Reds sent RHP A.J. Morris outright to Louisville Bats. Toronto Blue Jays recalled 2B Ryan Goins from Buffalo Bisons.
	2B Darwin Barney roster status changed by Toronto Blue Jays.
	Toronto Blue Jays optioned RHP Bo Schultz to Buffalo Bisons.
	New York Mets transferred CF Justin Ruggiano from the 15-day disabled list to the 60-day disabled list. Left shoulder injury.
	Los Angeles Angels traded RHP Fernando Salas to New York Mets for RHP Erik Manoah.
	New York Mets traded C Nevin Ashley to Texas Rangers for cash.
	Baltimore Orioles transferred RF Joey Rickard from the 15-day disabled list to the 60-day disabled list. Right thumb ligament injury.

Ingestion: pull down monthly JSON and store in AWS

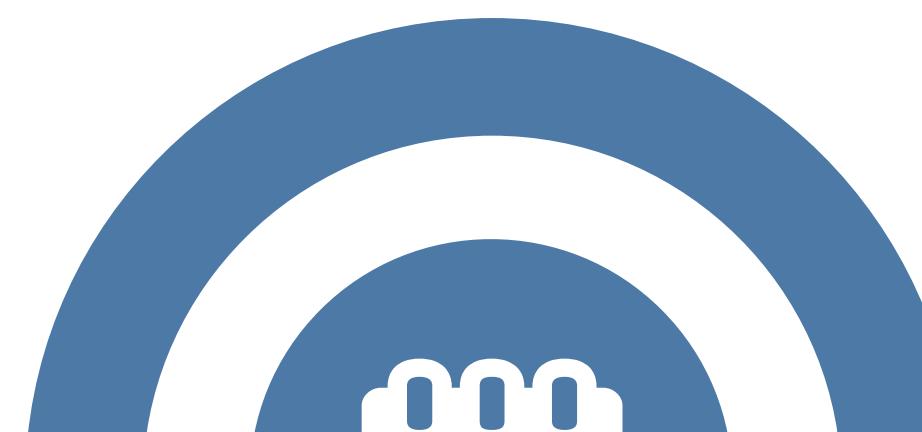
Also can pull nightly reports and add to correct monthly dump

Processing: Need to parse messages into useful data

JSON helped by providing dates, player and team ids and names

Regex to parse the data, built a mapping table to connect injury keywords (hamstring) to body parts (leg)

Future Improvements & Limitations



Integrate older transaction & play-by-play data

These have more difficult formats to parse



Better leverage PitchF/X

Make use of spin rate, ball movement, pitcher arm slot, etc.



Functions for Analysis

For example: comparing changes across multiple players with same injury, statistical significance of differences



Improve Interface

Improved search functions, and dynamic player pages



Limitations

- Limited to 2009 as earliest date because of JSON API
- Even if we could go back farther, PitchF/X only started in 2008