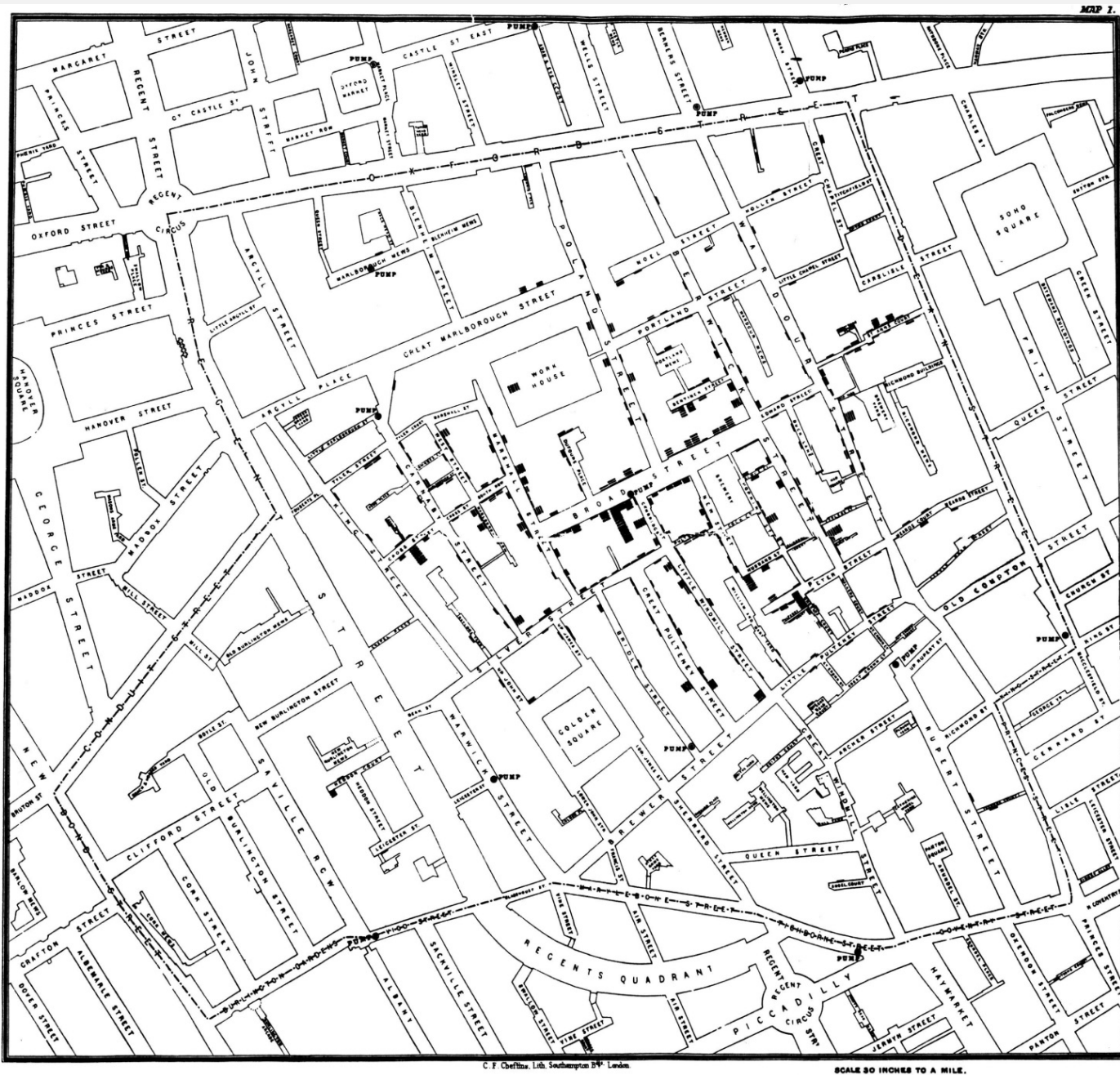


One BIG Additional Idea: Visualization

Telling Stories with Data

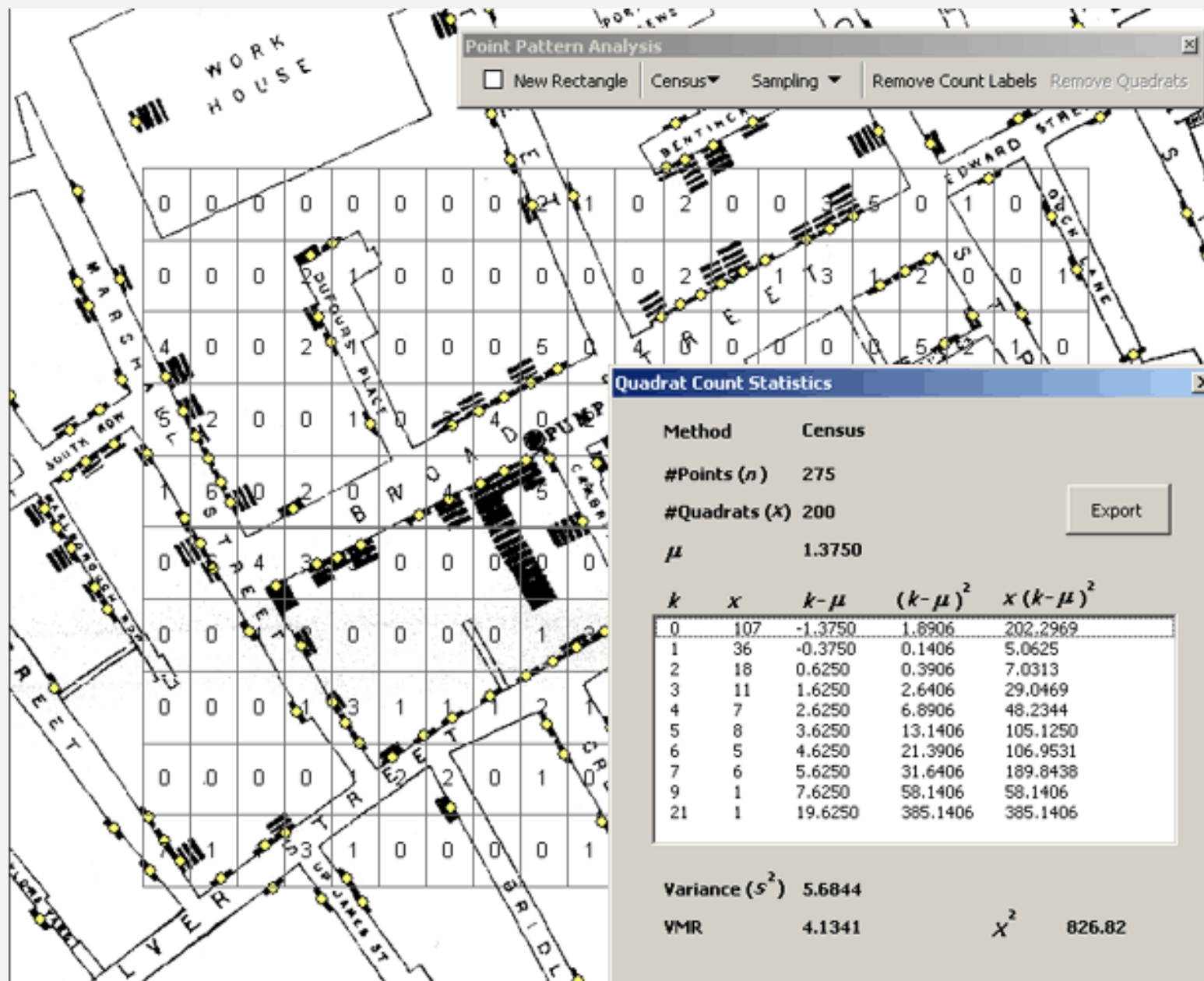


...where am I and what is missing in this picture?



Decoding the Outliers: John Snow's Cholera Map, 1854

**"Know There Are Unknowns: How suspected confounding variables influence models,"
I. Elaine Allen & Julia E. Seaman, Journal of Quality Progress, February, 2011*



Creating the grid to digitize John Snow's original map*.

*CartoDB & Digitizelt software: https://docs.cartodb.com/tutorials/conditional_styling/



With thanks to Andrew Hill

John Snow's Cholera Map, recreated from data extracted by the London Daily Examiner, 2013

What is unusual about the pattern of Cholera cases?



With thanks to Andrew Hill

John Snow's Cholera Map, recreated from data extracted by the London Daily Examiner, 2013

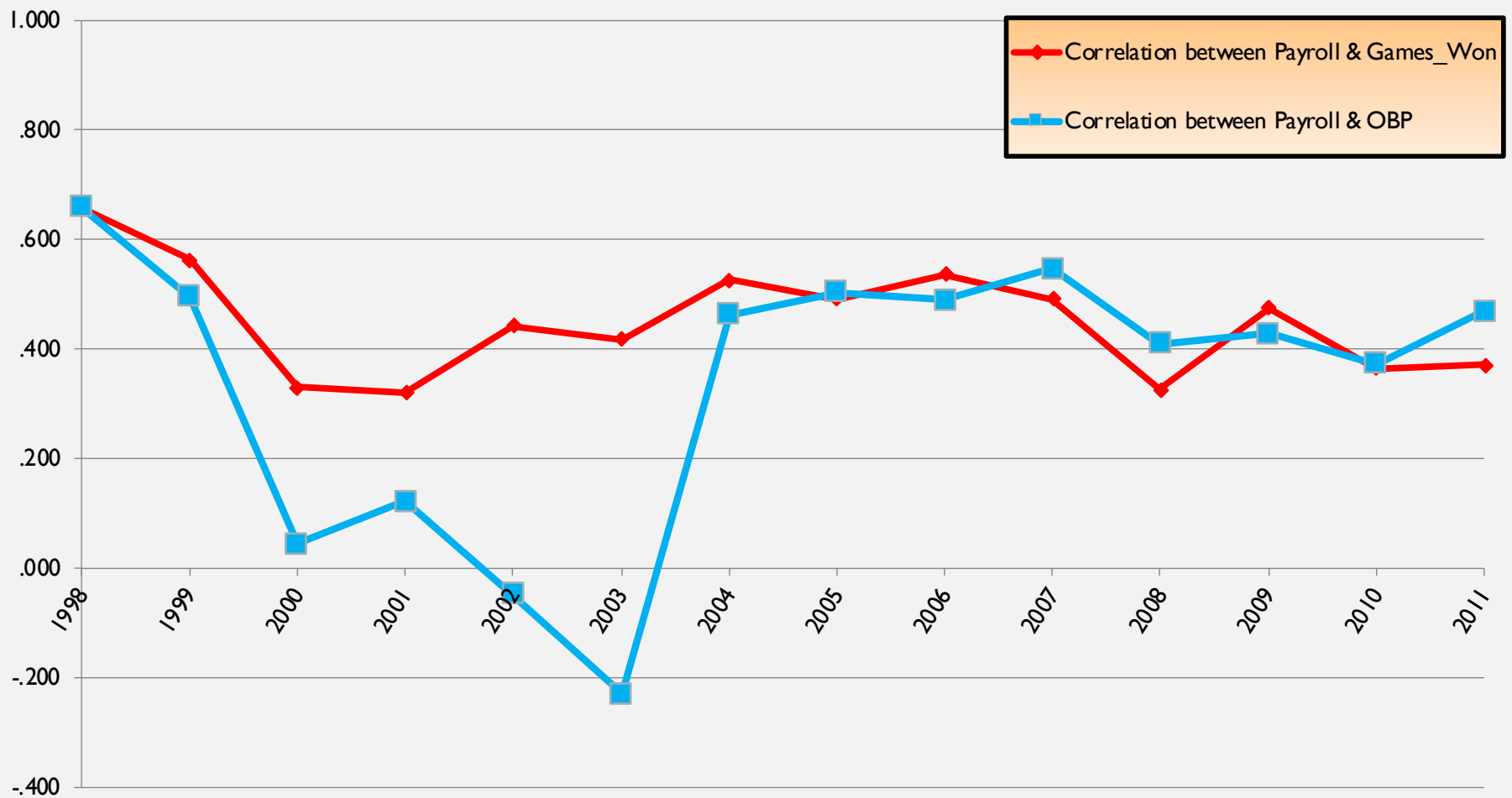
● The brewery had its own water source – the workers had water and ale

GRAPHICS TO TELL A STORY: OAKLAND A'S*

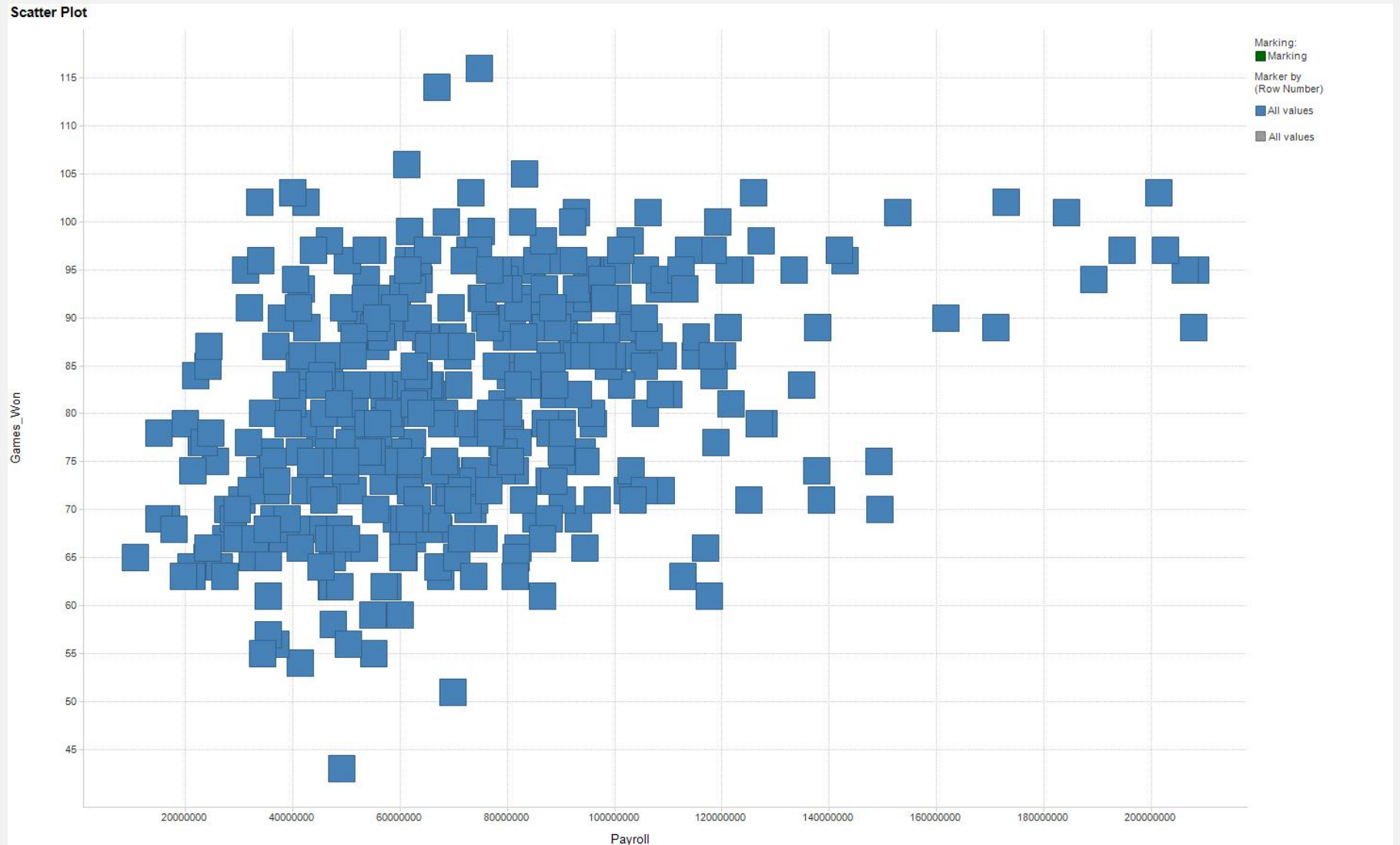
- 2004: *Moneyball* published by Michael Lewis
- 2011: The film 'Moneyball' released
 - The film ends with the A's winning their 20th consecutive game in 2002.
 - It's now 2019:
- Did Moneyball continue to work for the Oakland A's?
- Was Moneyball a sustainable practice?
- How far did Oakland go this year?

*"Fair or Foul? The innovative Moneyball management approach can make a difference—up to a point," Julia E. Seaman & I. Elaine Allen, *Journal of Quality Progress*, April, 2012.

MONEYBALL & THE OAKLAND A'S: BACKGROUND*

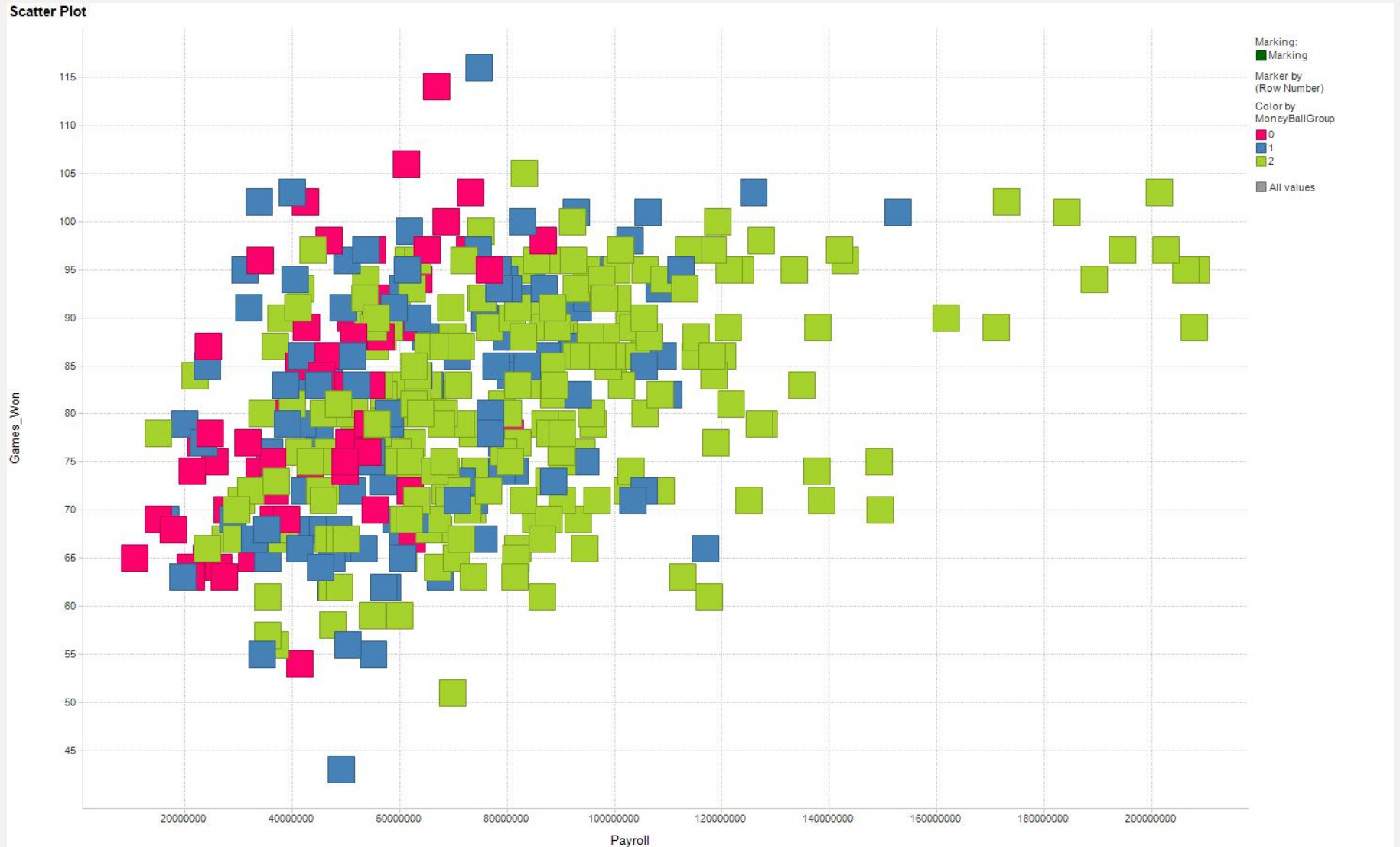


*Excel plot



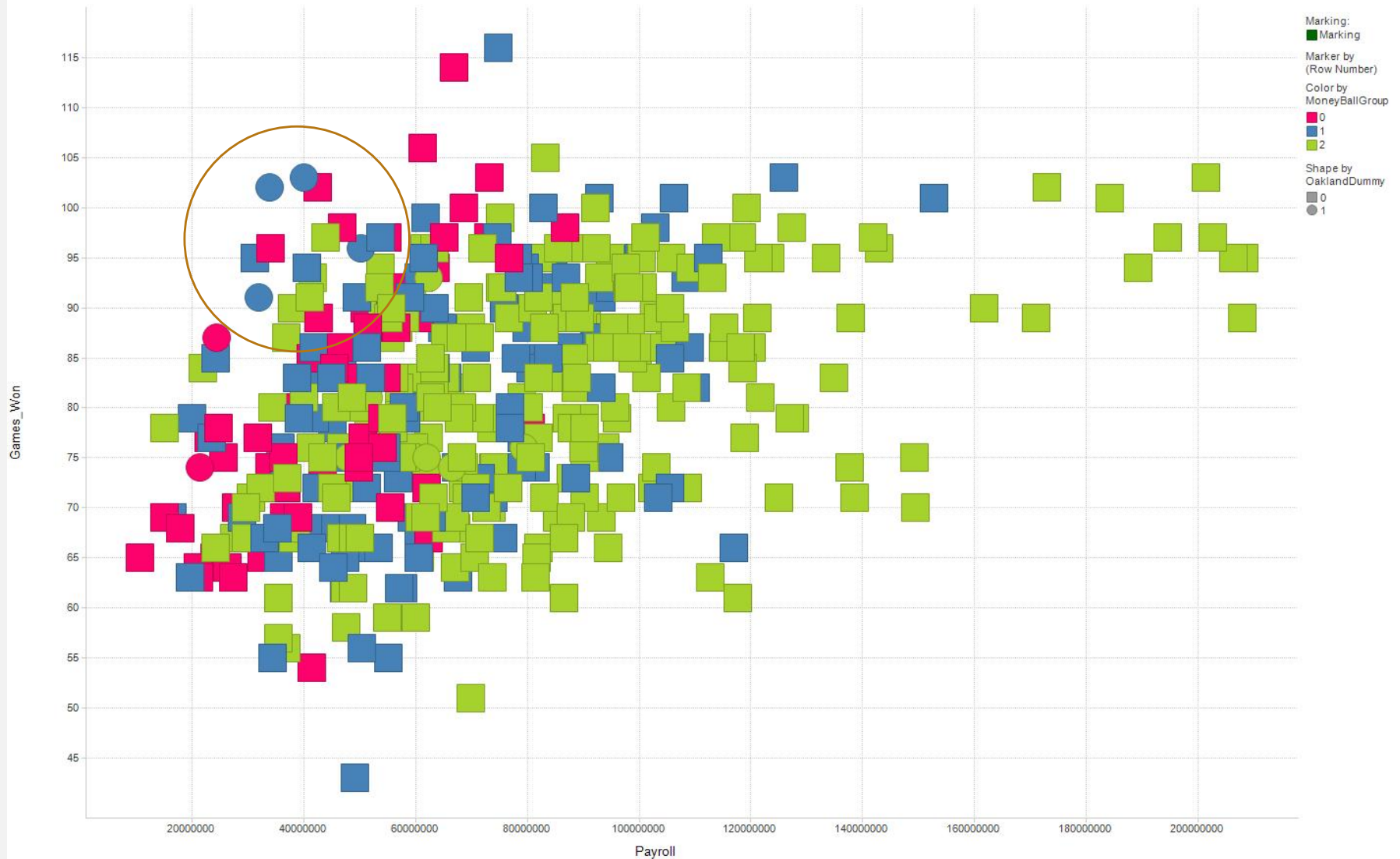
Payroll vs. Games Won*

*analysis & graphics completed with Excel & Spotfire

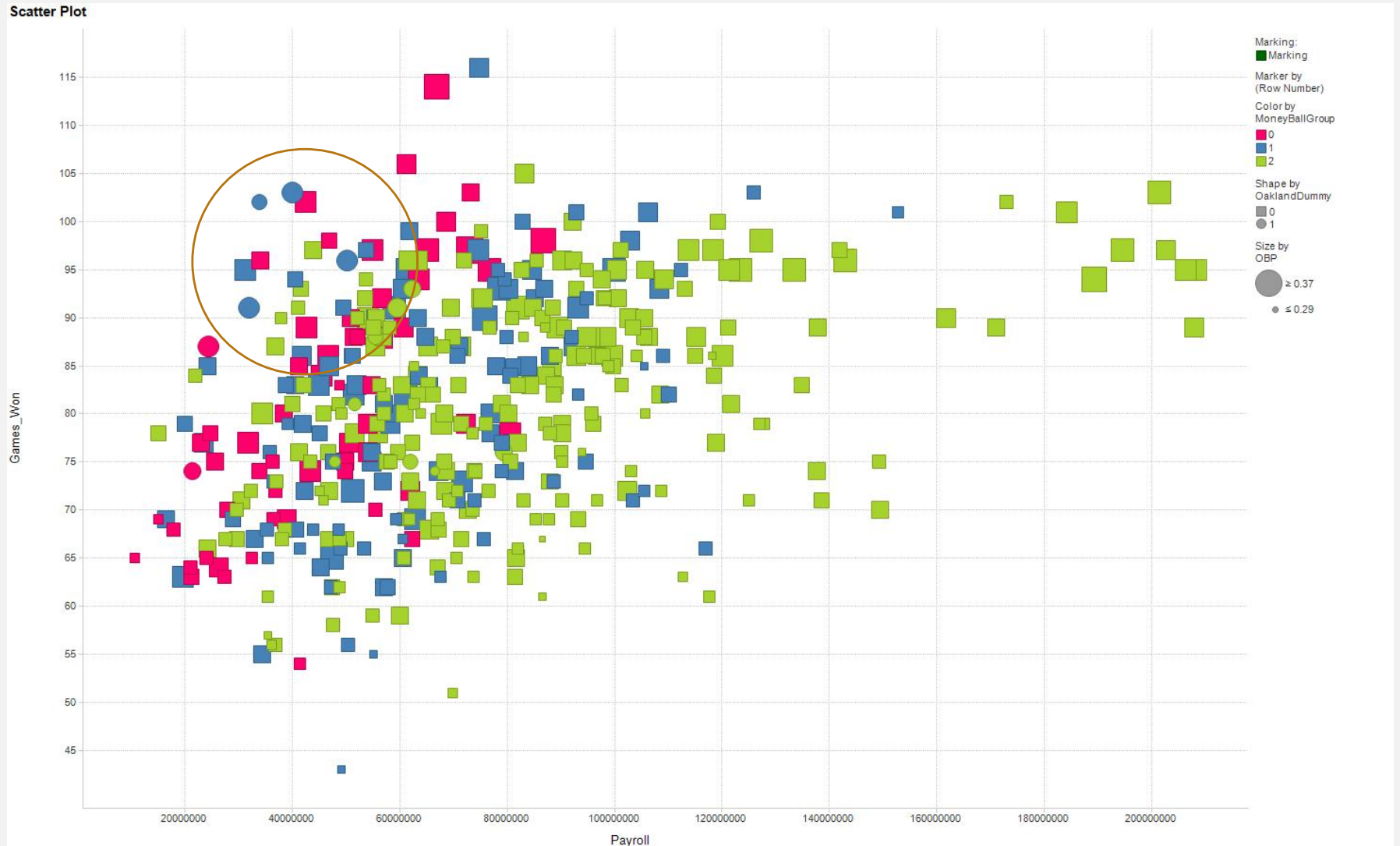


Overlay colors where **Red** = Pre-MoneyBall, **Blue** = MoneyBall, and **Green** is Post-MoneyBall

Scatter Plot

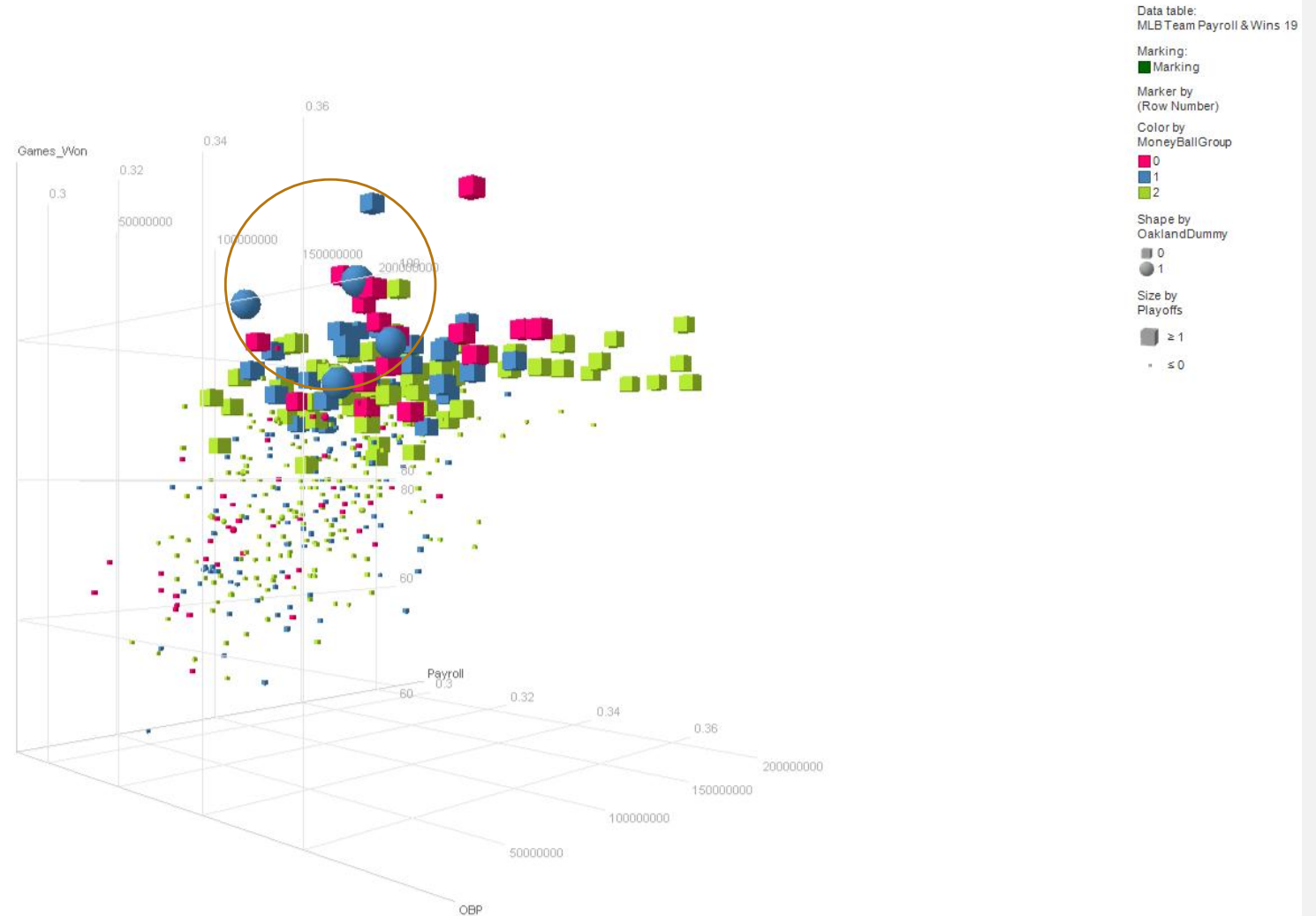


Add Circles for Oakland so **BLUE Circles** = Oakland & MoneyBall



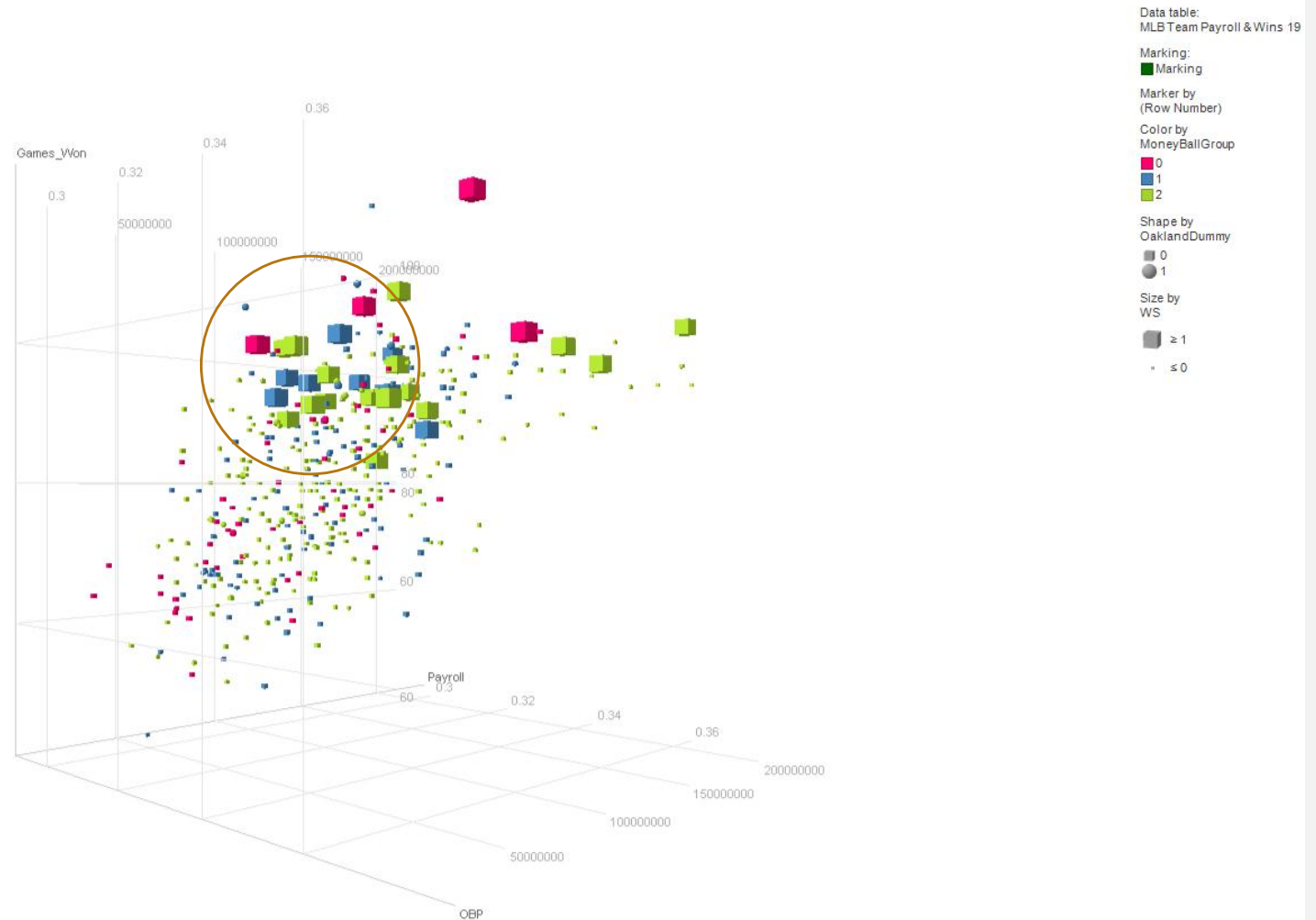
Add size to indicate the OBP for a team ... Oakland looking good!
But did it help them get to the Post Season?

3D Scatter Plot



Now in 3D – Look at Oakland during the MoneyBall era getting into the playoffs...but does using only one metric (OBP) = success?

3D Scatter Plot



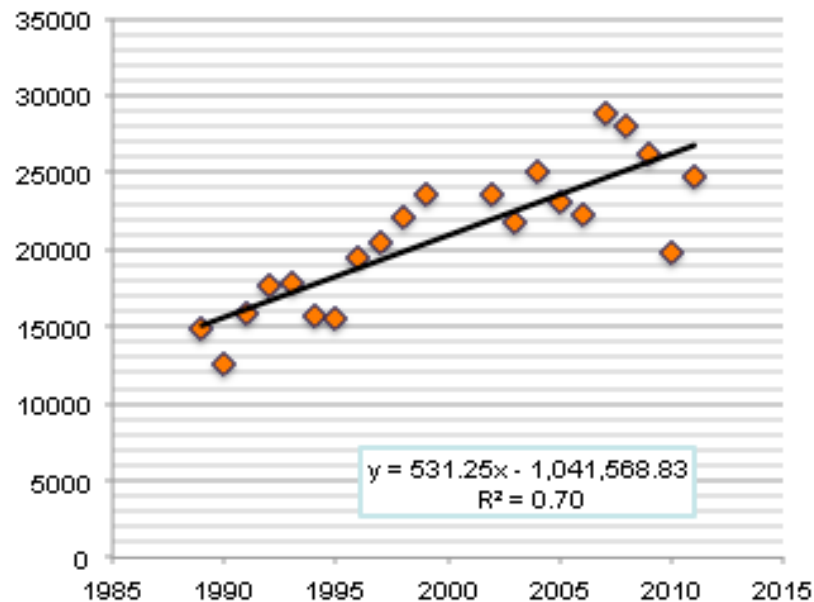
Sadly no, analytics can get you to the party but maximizing only one metric (offense) will not win you everything. In 2018 & 2019 Oakland made the playoffs but lost in the 1st round.

MLB & THE LA DODGERS: THE INJURY PROJECT

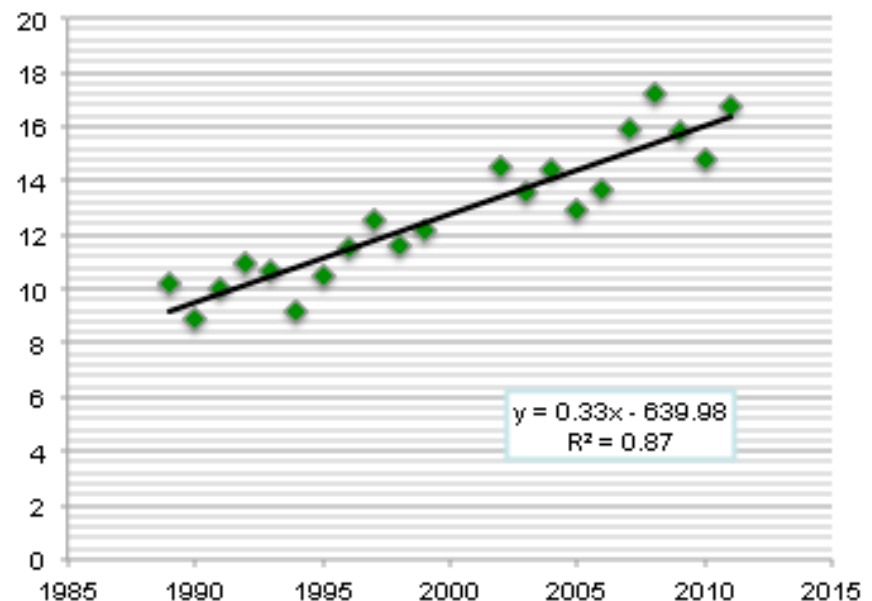
- Increasing number of players injured per year
- Increasing cost of injuries
- Negative correlation to training



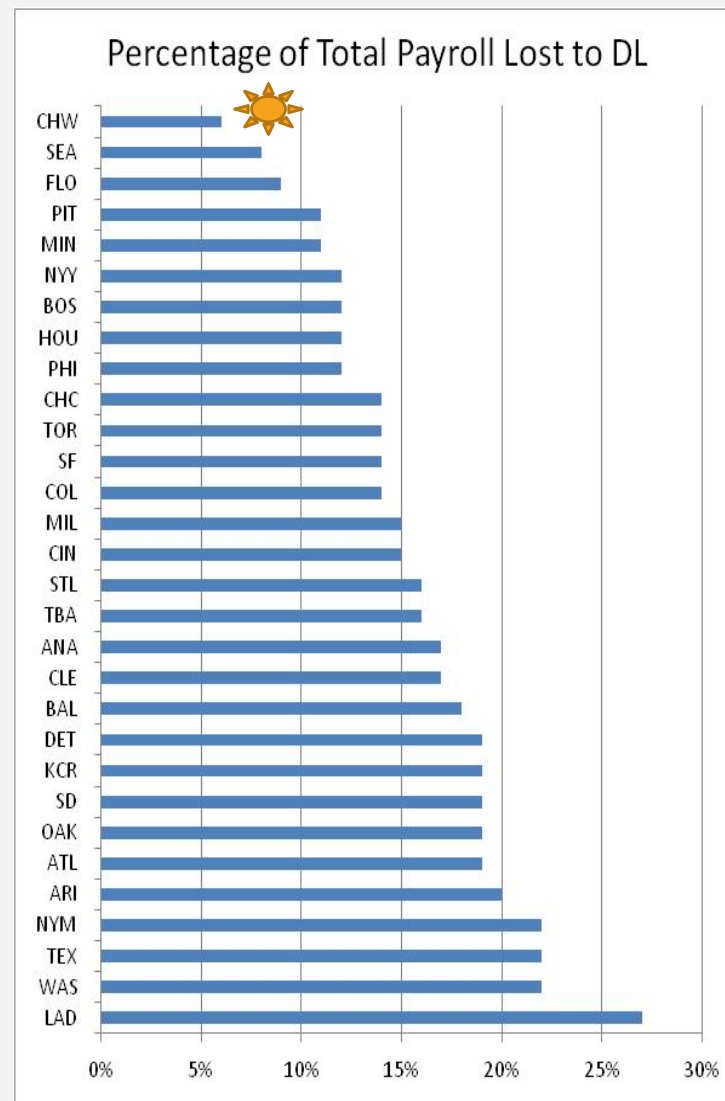
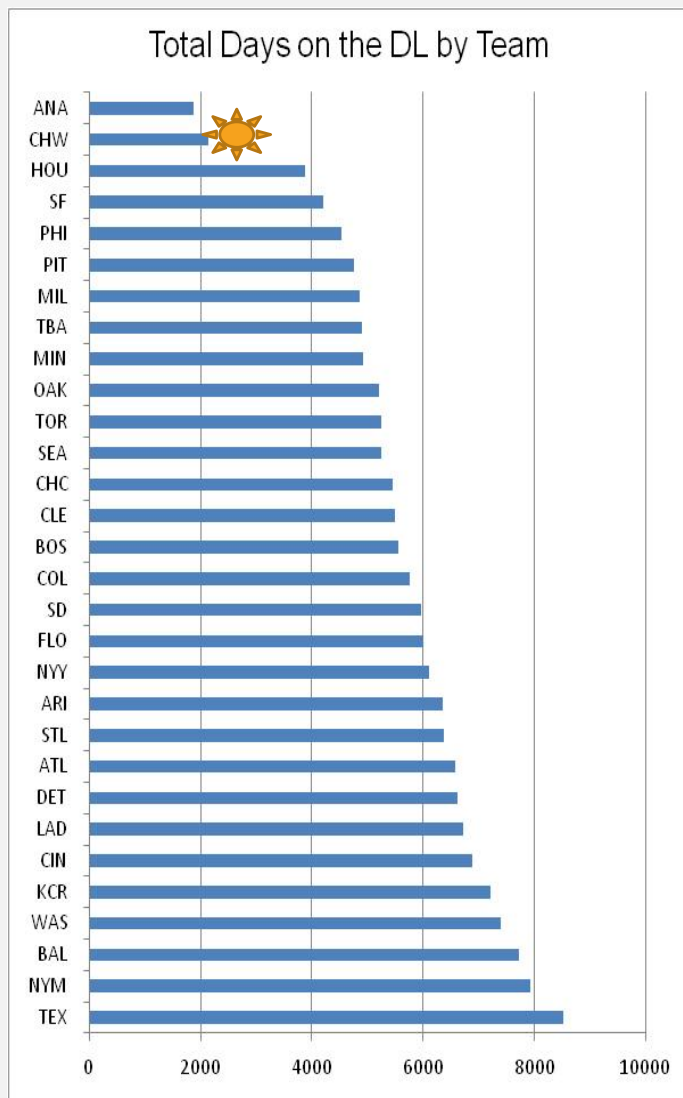
Total DL days by Year



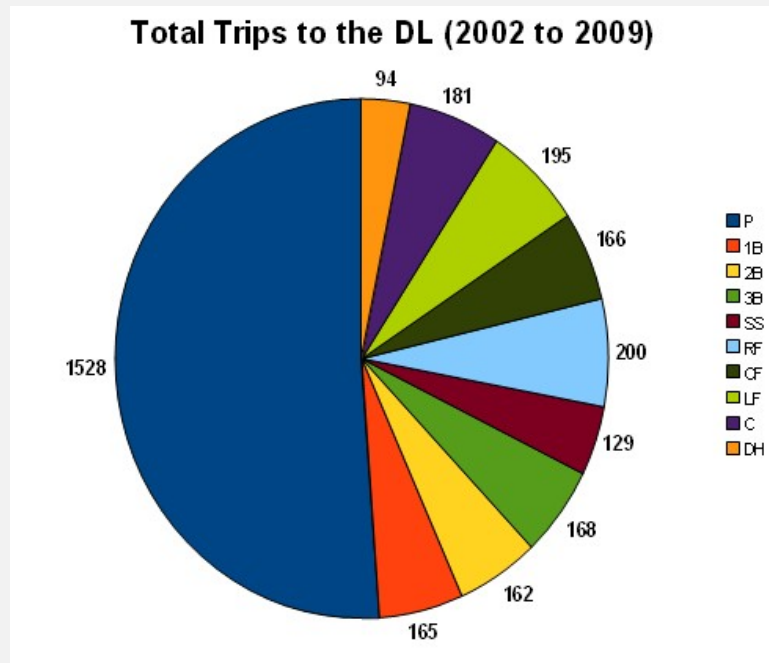
Average DL Trips per Team by Year



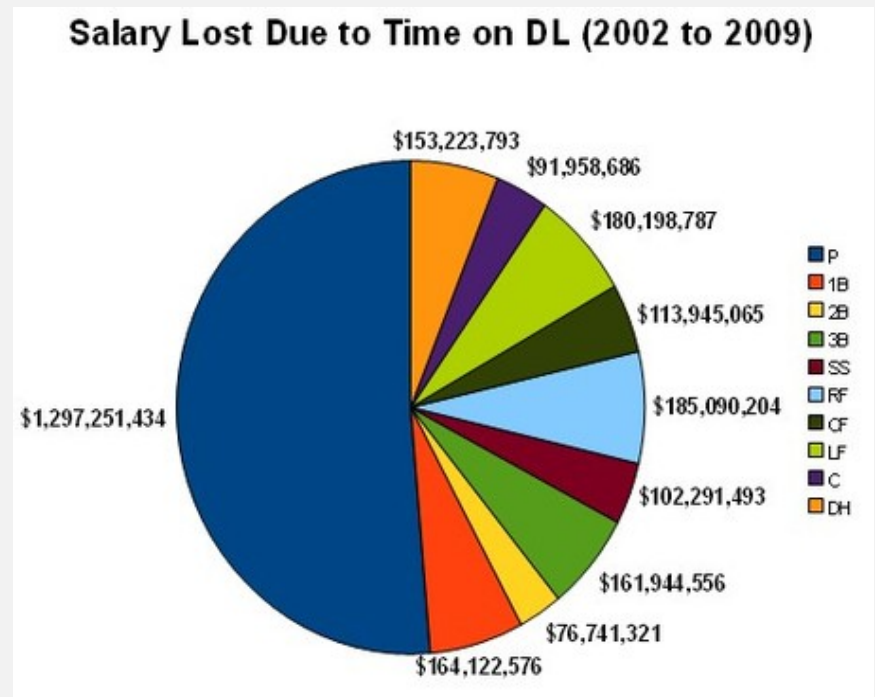
MLB & the LA Dodgers: The Injury Project



MLB & the LA Dodgers: The Injury Project



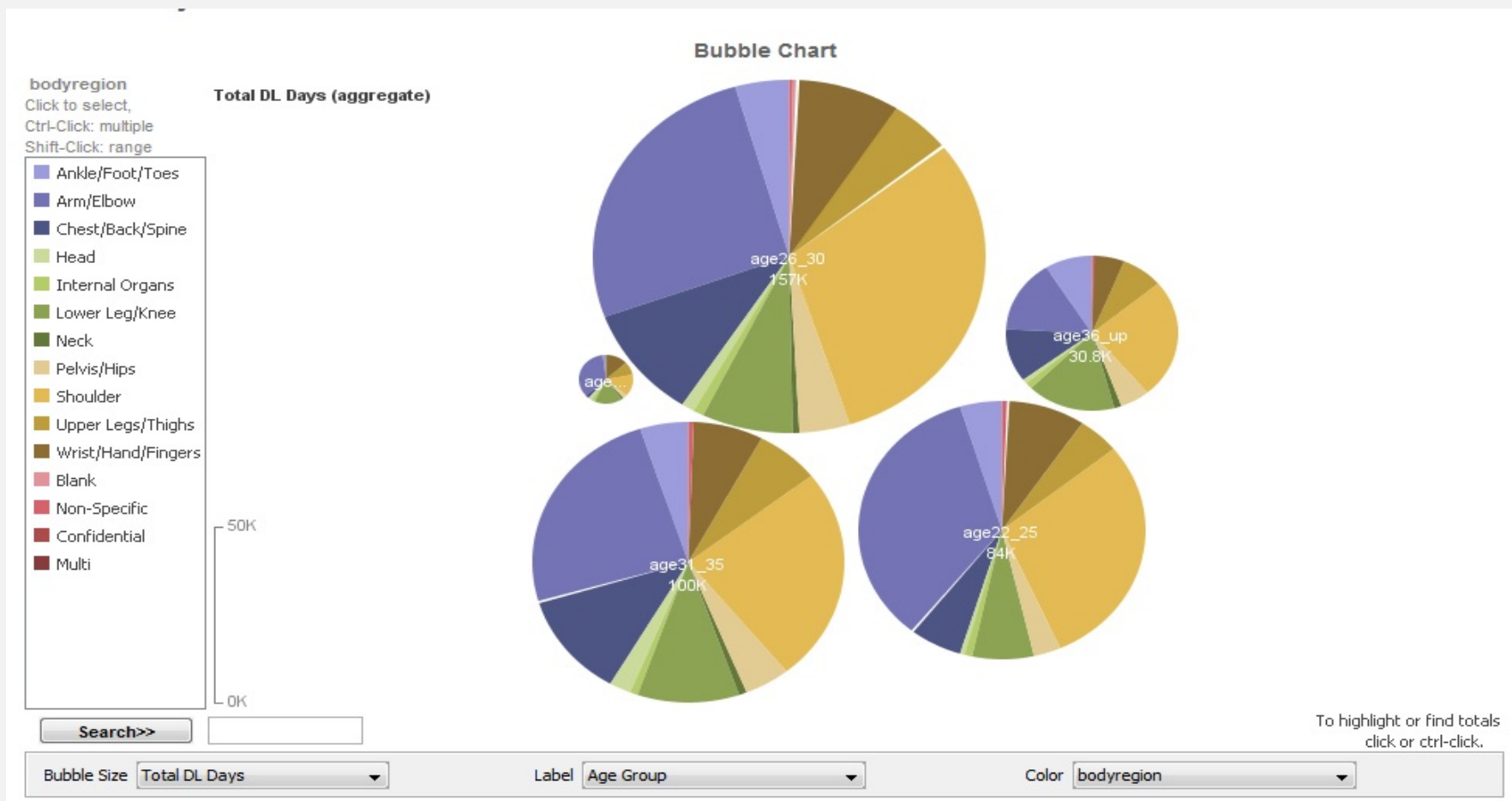
It's all about the pitchers



Need more detailed information about injuries by position ...

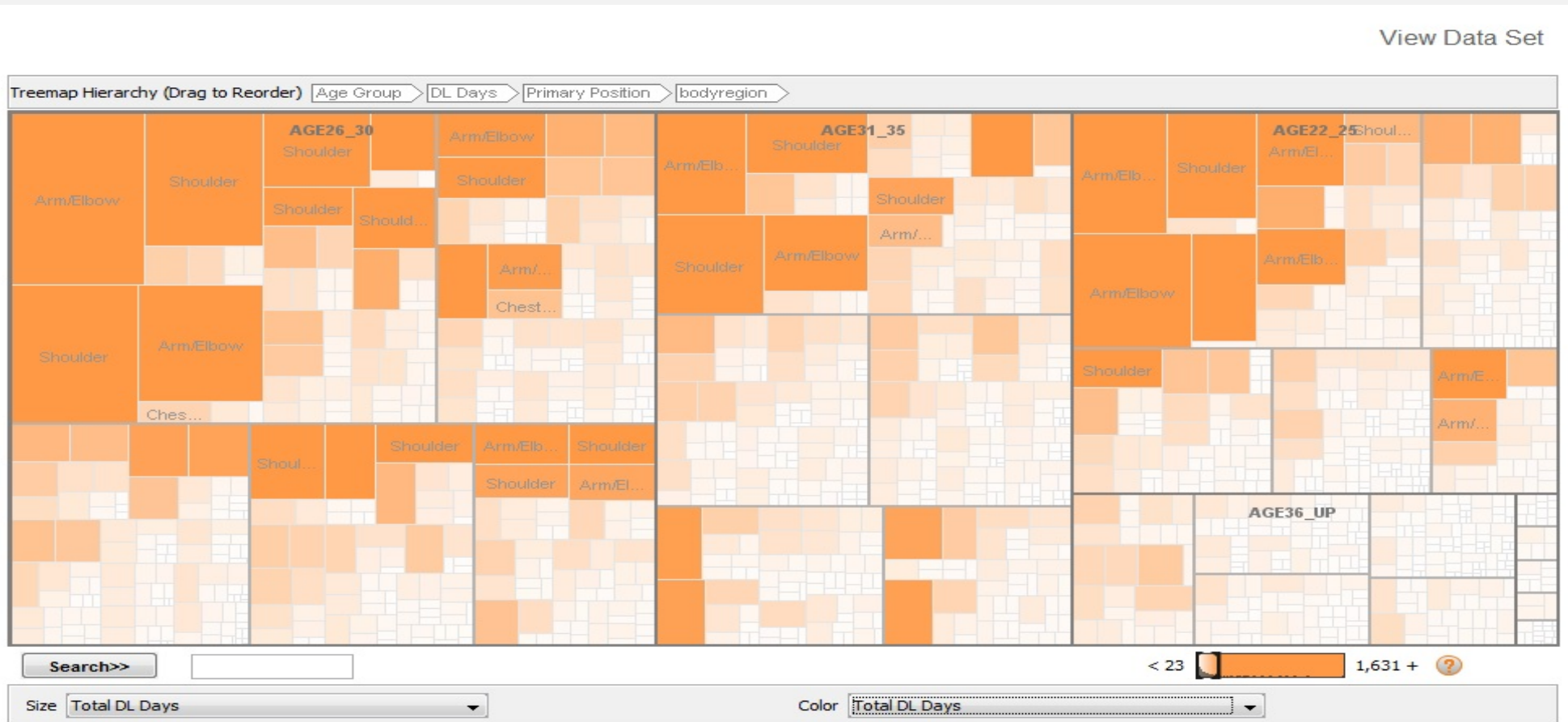
What about Age & Position:

- <http://www-958.ibm.com/software/analytics/labs/manyeyes/#home>
- <http://www-958.ibm.com/software/analytics/labs/manyeyes/#vis=501453>



What about Age & Position:

- <http://www-958.ibm.com/software/analytics/labs/manyeyes/#home>
- <http://www-958.ibm.com/software/analytics/labs/manyeyes/#vis=501453>



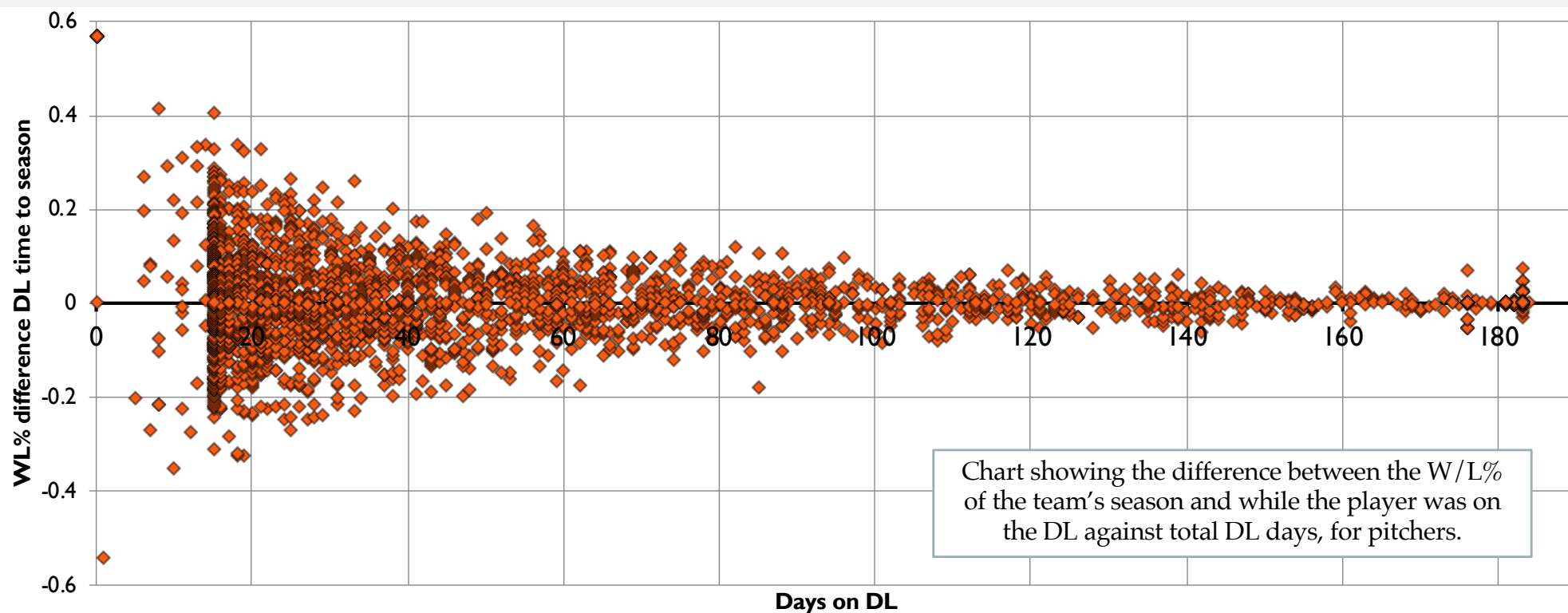
HOW TO ESTIMATE THE IMPACT OF BEING ON THE DL

- Examine the team performance pre & post the time that the player is on the DL (W/L ratio). Calculate the difference with the player on the field and on the DL*
- Example:
 - Damian Rolls, 2004 TB: Chest contusion
 - DL May 24- July 2, 39 days
 - TB's record during DL time: 25-10 (0.714)
 - TB's season 2004 record: 70-91 (0.435)
 - Rolls' DL W/L difference is -0.2795
 - **Tampa Bay winning percentage was almost 30% higher without Rolls**



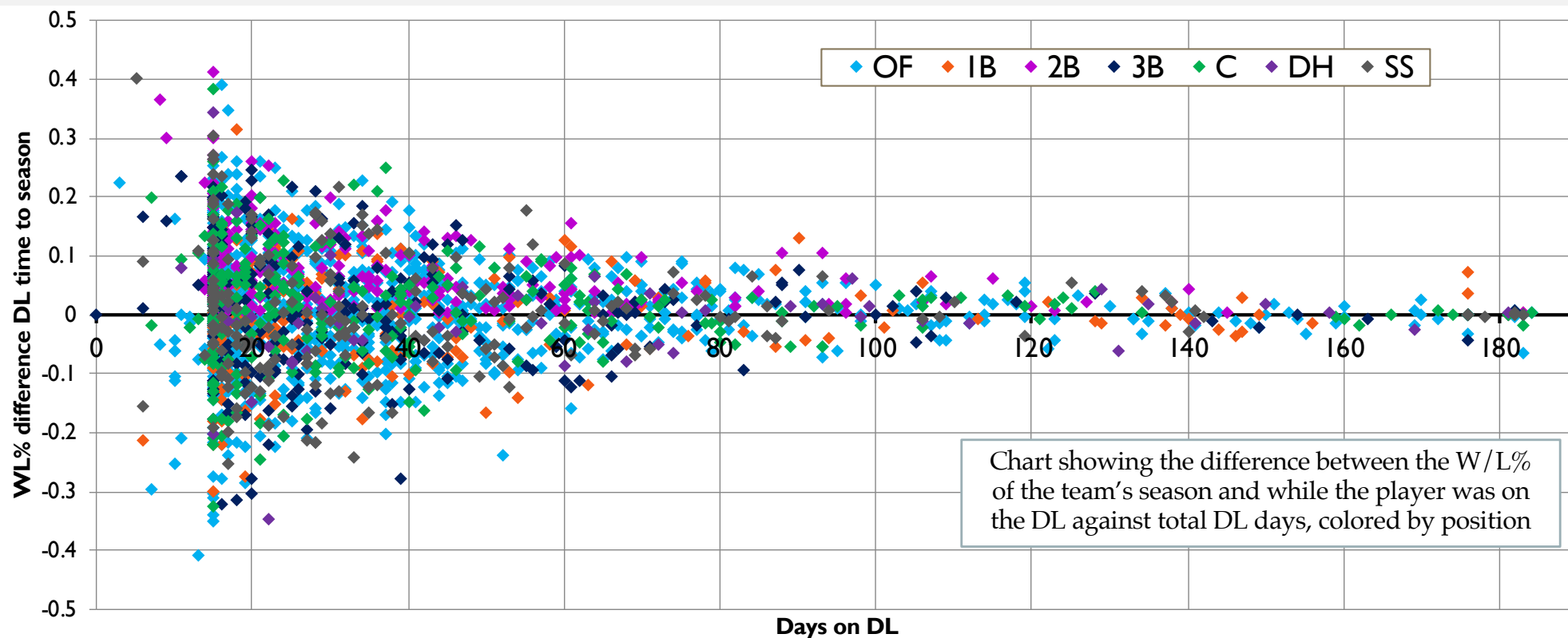
*"Better Training, Better Analytics, & More Players on the DL," Julia E. Seaman & I. Elaine Allen, Presented at the AmerStatAsso, Montreal, 2015

Pitchers



Rank	Player	Team	Year	Pos	Date on DL	Date off DL	Total Days	WL% Dif	Injury
1	Joey Devine	OAK	2011	P	27-Sep-2011	2-Nov-2011	1	-0.5432	rhomboid strain
2	Felix Doubront	BOS	2011	P	22-Mar-2011	8-Apr-2011	8	0.4127	forearm inflammation
3	Jon Lieber	PHI	2007	P	25-Mar-2007	9-Apr-2007	15	0.4065	Strained right oblique
4	Hector Ambriz	CLE	2010	P	23-Sep-2010	3-Oct-2010	10	-0.3519	Right elbow inflammation
5	Armando Benitez	SF	2006	P	14-Sep-2006	2-Oct-2006	18	0.3387	Right knee inflammation
6	Yorkis Perez	BAL	2002	P	16-Sep-2002	30-Sep-2002	14	0.3367	Appendectomy
7	Dennis Sarfate	BAL	2008	P	17-Sep-2008	30-Sep-2008	13	0.3315	Fractured clavicle
8	Joaquin Benoit	TEX	2003	P	1-Jun-2003	22-Jun-2003	21	0.3272	Tendinitis in right shoulder
9	Brad Penny	LAD	2005	P	5-Apr-2005	24-Apr-2005	19	-0.3264	Nerve irritation - right arm
10	Terry Mulholland	ARZ	2006	P	6-Jun-2006	21-Jun-2006	15	0.3263	Inner ear infection
11	Brad Lidge	PHI	2009	P	7-Jun-2009	26-Jun-2009	19	0.3241	Sprained right knee
12	Matt Ginter	CLE	2008	P	7-Aug-2008	25-Aug-2008	18	-0.3235	Strained right forearm
13	Dan Miceli	HOU	2004	P	22-Aug-2004	9-Sep-2004	18	-0.3210	Conjunctivitis
14	Mike Remlinger	CHC	2005	P	21-May-2005	5-Jun-2005	15	-0.3123	Fractured finger - left hand
15	Daniel Cabrera	BAL	2008	P	19-Sep-2008	30-Sep-2008	11	0.3112	Sprained pitching elbow

Position Players



Rank	Player	Team	Year	Pos	Date on DL	Date off DL	Total Days	WL% Dif	Injury
1	Denard Span	MIN	2011	OF	10-Jun-2011	23-Jun-2011	13	-0.4111	head concussion
2	Jeff Kent	SF	2002	2B	21-Mar-2002	6-Apr-2002	15	-0.4099	Fractured left wrist
3	Yuniesky Betancourt	KC	2009	SS	11-Jul-2009	16-Jul-2009	5	0.4012	Strained right hamstring
4	Jacque Jones	MIN	2003	OF	1-Jul-2003	17-Jul-2003	16	0.3889	Strained left groin
5	Chris Coste	PHI	2007	C	24-Mar-2007	8-Apr-2007	15	0.3827	Strained hamstring
6	Anderson Hernandez	WAS	2009	2B	5-Apr-2009	13-Apr-2009	8	0.3642	Strained left hamstring
7	Josh Willingham	FLA	2006	OF	7-Jun-2006	22-Jun-2006	15	-0.3519	Strained ligament - left hand
8	Jim Thome	MIN	2011	DH	2-Jun-2011	24-Jun-2011	22	-0.3480	quadriceps strain
9	Gary Matthews Jr.	BAL	2002	OF	25-Aug-2002	11-Sep-2002	17	0.3469	Right wrist tendinitis
10	Juan Encarnacion	LAD	2004	OF	4-Jul-2004	19-Jul-2004	15	-0.3426	Left shoulder tendinitis
11	Dean Palmer	DET	2002	DH	24-Mar-2002	7-Apr-2002	15	0.3416	Shoulder stiffness
12	Brandon Inge	DET	2002	C	12-May-2002	27-May-2002	15	-0.3251	Dislocated left shoulder
13	Kevin Youkilis	BOS	2004	3B	16-Aug-2004	1-Sep-2004	16	-0.3236	Bruised right ankle
14	Geoff Blum	SD	2005	3B	30-Apr-2005	18-May-2005	18	-0.3174	Left chest contusion
15	Clete Thomas	DET	2008	OF	4-Jun-2008	19-Jun-2008	15	-0.3124	Sprained right ankle

Team Effects?

Rank	Pos	Trips	Ave DifWL%
1	C	284	0.0110
2	3B	242	0.0070
3	2B	248	0.0037
4	SS	206	0.0031
5	OF	804	0.0029
6	P	2443	0.0021
7	DH	60	-0.0023
8	IB	198	-0.0046

- Overall, teams did slightly worse while players were on DL
 - Mean DifWL%= 0.003 (not significantly different)
 - Teams lost most without catchers and won most without 1st Basemen.
- Infield players hurt a team more than outfielders

This is a summary table of linear regressions of total DL Days and season WL%, 2002-2011 for each team. A few teams show a larger direct correlation (always negative) but most teams are very loosely affected.

Team	Direction	R2
ARZ	-	0.16568
ATL	-	0.54194
BAL	+	0.01572
BOS	-	0.0226
CHC	+	0.00052
CIN	-	0.73704
CLE	-	0.13732
COL	-	0.00114
CWS	-	0.20728
DET	-	0.00667
FLA	+	0.00113
HOU	-	0.63405
KC	-	0.18294
LAA	+	0.05581
LAD	-	0.25082
MIN	-	0.16924
MLW	-	0.06054
NYM	+	0.0149
NYN	-	0.18393
OAK	-	0.64791
PIT	+	0.04208
SD	-	0.57971
SEA	-	0.00057
SF	-	0.20632
STL	-	0.32249
TB	-	0.08225
PHI	+	0.24191
TEX	-	0.01923
TOR	-	0.08326
WAS	-	0.04226

Motivating Example: UCSF Project Studying Pedestrians Hit by Cars in San Francisco

- Adjusting the traffic-related injury by age, gender, occupation, # of dependents and ethnicity. Map the injuries and use DALYs for San Francisco for comparisons to global, regional, state, and city trends.
- Overall Mean Disability weight for an ankle sprain = 0.064 YLD or 14 days
- A 38 year old male with a life expectancy of 77 years who sprains his ankle has a DALY of 0.03 years or 11 days
- A 66 year old male with the same injury has a DALY of 0.06 years or 22 days
- **Question for SF: Where are traffic lights needed?**

SHORT QUIZ on San Francisco auto/pedestrian accidents:

- When do most pedestrians get hit by cars?
- What neighborhood or area are they most likely to get hit?
- What is the most likely confounding variable in getting hit by a car?
- What gender and age group(s) are most likely hit by a car?



Creating Disability Adjusted Player Days (DAPD)

Disability Adjusted Player Days (DAPD) adjusted for:

- Average time lost for an injury
- Previous injuries
- Position the player plays
- Age of the player
- Expected future playing years in the MLB
- Years in position in MLB
- Years in a contract
- Team you play for
- Others?



Example: Creating the DAPD for a Catcher with an Ankle Sprain

$$\text{DAPD} = - \left[\frac{DCe^{-\beta a}}{(\beta+r)^2} \right] \left[e^{-(\beta+r)(L)} (1 + (\beta+r)(L+a)) - (1 + (\beta+r)a) \right]$$

- L = Adjust for average time lost for an injury
 - mean = 44 days, median = 40 days [15 to 140]*
- C = age of the player = 33 years old in this example
- r = Standard discounted lifespan of 3% per year
- DW = Disability weight for an ankle sprain = 0.064

Calculate β for adjustment factors:

- K_1 = Adjust for expected future years the player will play this catcher reached the MLB at age 25, & probably has at most 4 more active years
- K_2 = Adjust for time already played in MLB = 8 years
- K_3 = Adjust for Catcher DL days mean = 49 days, median = 29 days

*based on 14 years of DL data including MLB EMR data

Example: Creating the DAPD for a Catcher with an Ankle Sprain

- Calculating the effect of K_1 , K_2 & K_2 separately for Catchers with ankle sprains:
- No adjustment = **55** days
- K_1 = **59** days on the DL
- K_2 = **60** days on the DL
- K_3 = **61** days on the DL
- Unclear on the best way combine K_1 , K_2 & K_2 to create β for the adjustment:

Example: Sensitivity of Adjustments

- Given you are a catcher, as age changes from 25 to 35, the DL days for an ankle sprain increase by .25 per year

AGE	25	26	27	28	29	30	31	32	33	34	35
DL Days	45.95	46.18	46.39	46.60	46.80	46.99	47.16	47.33	47.48	47.63	47.77

- Given you are 35 years old, as position changes, DL days change:

Position	1B	2B	3B	C	DH	LHR	LHS	OF	RHR	RHS	SS
DL Days	51	52	53	57	66	71	80	60	80	86	71

- Given you are a Catcher, as time in the majors changes (as % of overall expected playing time):

Fraction of expected MLB time	90%	85%	80%	75%	70%	65%	60%	55%	50%
DL Days	70.96	67.88	65.34	63.22	61.42	59.89	58.57	57.43	56.43

(DL years 1998 – 2015):

- Only one player was on the DL because of a bullet wound
- Only one player was on the DL because of insomnia
- Troy O’Leary, Russ Johnson, & Trip Cromer are the only 3 players to go on the DL for this.