

# Analytics in Sports: A Brief History

# Analytics in Sports:

## A Brief History

BOSTON.							ATHLETIC.																	
G. Wright, s.s.	T.	R.	l.	b.	P.	O.	A.	E.	Force, s. s.	T.	R.	l.	b.	P.	O.	A.	E.							
Leonard, 2b.	6	4	4	1	5	2			Eggler, c. f. <td>5</td> <td>3</td> <td>3</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td></td>	5	3	3	0	0	0	0								
O'Rourke, 1b	6	3	3	4	4	3			Fisler, r. f. <td>5</td> <td>0</td> <td>1</td> <td>2</td> <td>0</td> <td>0</td> <td>0</td> <td></td>	5	0	1	2	0	0	0								
Murnan, l. f.	6	1	0	3	1	0			Meyerle, 3db	5	1	2	2	3	3									
Schafer, 2d b.	6	3	3	3	1	2			Sutton, 1st b.	5	1	2	10	0	0									
McGinley, c.f	6	0	0	0	0	1			Coons, c.	5	1	0	1	1	3									
Manning, r.f.	6	0	2	2	0	0			Hall, l. f.	5	1	3	5	0	0									
Morrill, c.	6	2	2	4	1	2			Fowler, 2d b.	6	1	2	6	7	5									
Josephs, p.	5	4	4	1	1	2			Knight, p.	5	2	2	0	1	2									
Totals... 53							19	21	27	13	13	Totals... 46							11	17	27	15	15	
Boston.....							0	1	3	3	4	1	Athletic.....							0	2	2	0	11
Athletic.....							1	0	0	0	3	3	Home-run—Hall, l.							2	2	0	1	
Runs earned—Boston, 4; Athletic, 5. Home-run—Hall, l.																								
Total bases on hits—Boson, 22; Athletic, 20. First base by																								
errors—Boston, 8; Athletic, 5. Umpire, George White of																								
Lowell, Mass. Time 2h. 47m.																								

Then

# Analytics in Sports: A Brief History

BOSTON.						ATHLETIC.																	
G.	T.	R.	B.	P.	O.	A.	E.	G.	T.	R.	B.	P.	O.	A.	E.								
G. Wright, s.s.	6	4	4	1	5	2		Force, s.s.	5	1	2	1	3	2									
Leonard, 2b.	6	3	3	4	4	3		Eggler, c. f.	5	3	3	0	0	0									
O'Rourke, 1b.	6	2	3	9	0	1		Fisler, r. f.	5	0	1	2	0	0									
Murnan, l. f.	6	1	0	3	1	0		Meyerle, 3db.	5	1	2	2	3	3									
Schafer, 2d b.	6	3	3	3	1	2		Sutton, 1st b.	5	1	2	10	0	0									
McGinley, c. f.	6	0	0	0	0	1		Coons, c.	5	1	0	1	1	3									
Manning, r. f.	6	0	2	2	0	0		Hall, l. f.	5	1	3	5	0	0									
Morrill, c.	6	2	2	4	1	2		Fowler, 2d b.	6	1	2	6	7	5									
Josephs, p.	5	4	4	1	1	2		Knight, p.	5	2	2	0	1	2									
Totals.....						53	19	21	27	13	13	Totals.....						46	11	17	27	15	15
Boston.....						0	1	3	3	4	1	Boston.....						0	2	5	19		
Athletic.....						1	0	0	0	3	3	Athletic.....						2	2	0	11		
Runs earned—Boston, 4; Athletic, 5. Home-run—Hall, l.																							
Total bases on hits—Boston, 22; Athletic, 20. First base by																							
errors—Boston, 8; Athletic, 5. Umpire, George White of																							
Lowell, Mass. Time 2h. 47m.																							

Then



Now

# What are sports analytics?

- **Sports analytics**, simply stated, is the quantification and analysis of on-field or on-court outcomes in sports using data

# What are sports analytics?

- **Sports analytics**, simply stated, is the quantification and analysis of on-field or on-court outcomes in sports using data
- Analytics in its simplest form predates professional sports itself

# What are sports analytics?

- **Sports analytics**, simply stated, is the quantification and analysis of on-field or on-court outcomes in sports using data
- Analytics in its simplest form predates professional sports itself
- The simplest metrics give fans a way to understand the outcomes of games they can't watch themselves

# What are sports analytics?

- **Sports analytics**, simply stated, is the quantification and analysis of on-field or on-court outcomes in sports using data
- Analytics in its simplest form predates professional sports itself
- The simplest metrics give fans a way to understand the outcomes of games they can't watch themselves
- Nowadays, measuring the actual and predicted contribution of players to on-field success is big business

# What are sports analytics?

- **Sports analytics**, simply stated, is the quantification and analysis of on-field or on-court outcomes in sports using data
- Analytics in its simplest form predates professional sports itself
- The simplest metrics give fans a way to understand the outcomes of games they can't watch themselves
- Nowadays, measuring the actual and predicted contribution of players to on-field success is big business
- Nearly all North American professional teams have an analytics department; the rise of modern sports analytics was famously profiled in the 2003 book and subsequent movie **Moneyball**



# Analytics and Gambling

- More general interest in sports analysis has increased since the US Supreme Court legalized sports gambling in 2018

# Analytics and Gambling

- More general interest in sports analysis has increased since the US Supreme Court legalized sports gambling in 2018
  - *Murphy v. National Collegiate Athletic Association*

# Analytics and Gambling

- More general interest in sports analysis has increased since the US Supreme Court legalized sports gambling in 2018
  - *Murphy v. National Collegiate Athletic Association*
- Canadian law has traditionally permitted some sports gambling, although the modern sports betting industry has exploded with the rise of internet betting sites

# Analytics and Gambling

- More general interest in sports analysis has increased since the US Supreme Court legalized sports gambling in 2018
  - *Murphy v. National Collegiate Athletic Association*
- Canadian law has traditionally permitted some sports gambling, although the modern sports betting industry has exploded with the rise of internet betting sites
- 52% of Canadians 18-34 are estimated to have placed an online sports bet at some point as of 2018

# Analytics and Gambling

- More general interest in sports analysis has increased since the US Supreme Court legalized sports gambling in 2018
  - *Murphy v. National Collegiate Athletic Association*
- Canadian law has traditionally permitted some sports gambling, although the modern sports betting industry has exploded with the rise of internet betting sites
- 52% of Canadians 18-34 are estimated to have placed an online sports bet at some point as of 2018
- Today the sports analytics community is quite large, including a lot of amateurs, and based on a wide variety of public, semi-public and proprietary data

# Analytics and Gambling

- More general interest in sports analysis has increased since the US Supreme Court legalized sports gambling in 2018
  - *Murphy v. National Collegiate Athletic Association*
- Canadian law has traditionally permitted some sports gambling, although the modern sports betting industry has exploded with the rise of internet betting sites
- 52% of Canadians 18-34 are estimated to have placed an online sports bet at some point as of 2018
- Today the sports analytics community is quite large, including a lot of amateurs, and based on a wide variety of public, semi-public and proprietary data
- Analytics has also come to nearly all sports, with soccer seeing the most recent uptick in analytical interest

# The beginning of analytics: Baseball

- Among the major North American sports, the idea of quantifying on-field impact began (naturally) with baseball

# The beginning of analytics: Baseball

- Among the major North American sports, the idea of quantifying on-field impact began (naturally) with baseball
  - The first box score appeared in a New York newspaper in 1845
  - The idea of separating a player's contribution from his team's performance, and the best way to do so, was already a hot topic of discussion among aficionados by the 1850s



# The beginning of analytics: Baseball

- Among the major North American sports, the idea of quantifying on-field impact began (naturally) with baseball
  - The first box score appeared in a New York newspaper in 1845
  - The idea of separating a player's contribution from his team's performance, and the best way to do so, was already a hot topic of discussion among aficionados by the 1850s
- Baseball being the origin sport for the development of sports analytics is largely due to baseball's pedigree

# The beginning of analytics: Baseball

- Among the major North American sports, the idea of quantifying on-field impact began (naturally) with baseball
  - The first box score appeared in a New York newspaper in 1845
  - The idea of separating a player's contribution from his team's performance, and the best way to do so, was already a hot topic of discussion among aficionados by the 1850s
- Baseball being the origin sport for the development of sports analytics is largely due to baseball's pedigree
- However, it is also due to the fact that separating player from team performance is relatively easy

# The beginning of analytics: Baseball

- Among the major North American sports, the idea of quantifying on-field impact began (naturally) with baseball
  - The first box score appeared in a New York newspaper in 1845
  - The idea of separating a player's contribution from his team's performance, and the best way to do so, was already a hot topic of discussion among aficionados by the 1850s
- Baseball being the origin sport for the development of sports analytics is largely due to baseball's pedigree
- However, it is also due to the fact that separating player from team performance is relatively easy
- This is because, in baseball compared to other sports, the action involves relatively little coordination between teammates

# Measuring player performance in baseball

- Baseball analytics was pioneered by Henry Chadwick (1824-1908), the so-called (and self-called) “Father of Baseball”

# Measuring player performance in baseball

- Baseball analytics was pioneered by Henry Chadwick (1824-1908), the so-called (and self-called) “Father of Baseball”
- Chadwick came from a prominent family of British intellectuals (mostly musicians and scientists) who immigrated to Brooklyn in 1836, adopting a career as a baseball beat writer and promoter

# Measuring player performance in baseball

- Baseball analytics was pioneered by Henry Chadwick (1824-1908), the so-called (and self-called) “Father of Baseball”
- Chadwick came from a prominent family of British intellectuals (mostly musicians and scientists) who immigrated to Brooklyn in 1836, adopting a career as a baseball beat writer and promoter
- Among Chadwick’s contributions were the development of the modern box score, with numbers indicating player positions and a “K” to record a strike out

# Measuring player performance in baseball

- Baseball analytics was pioneered by Henry Chadwick (1824-1908), the so-called (and self-called) “Father of Baseball”
- Chadwick came from a prominent family of British intellectuals (mostly musicians and scientists) who immigrated to Brooklyn in 1836, adopting a career as a baseball beat writer and promoter
- Among Chadwick’s contributions were the development of the modern box score, with numbers indicating player positions and a “K” to record a strike out
- He produced the first baseball guide, first called *The Beadle Dime Base-Ball Player* and later edited the *Spalding Base Ball Guides*, in collaboration with his friend A.G. Spalding (of sporting goods fame)

# Chadwick's stats

- Chadwick emphasized the development of statistics to allow fans to assess baseball players' performance in an *objective* manner:

*In order to obtain an accurate estimate of a player's skill, an analysis, both of his play at the bat and in the field, should be made, inclusive of the way in which he was put out; and that this may be done, it is requisite that all...contests should be recorded in a uniform manner.*

– Chadwick (1861)



# Chadwick's stats

- Chadwick emphasized the development of statistics to allow fans to assess baseball players' performance in an *objective* manner:

*In order to obtain an accurate estimate of a player's skill, an analysis, both of his play at the bat and in the field, should be made, inclusive of the way in which he was put out; and that this may be done, it is requisite that all...contests should be recorded in a uniform manner.*

– Chadwick (1861)

- To this end, Chadwick popularized two baseball impact statistics still used today:

# Chadwick's stats

- Chadwick emphasized the development of statistics to allow fans to assess baseball players' performance in an *objective* manner:

*In order to obtain an accurate estimate of a player's skill, an analysis, both of his play at the bat and in the field, should be made, inclusive of the way in which he was put out; and that this may be done, it is requisite that all...contests should be recorded in a uniform manner.*

– Chadwick (1861)

- To this end, Chadwick popularized two baseball impact statistics still used today:
  1. **Batting average**: hits by a player divided by his at-bats
  2. **Earned run average** (ERA): runs per 9 innings surrendered by a pitcher not due to fielding errors

# Batting average and its discontents

- Batting average is still widely cited as a measure of a hitter's quality today

# Batting average and its discontents

- Batting average is still widely cited as a measure of a hitter's quality today
- However, it quickly developed some critics!

# Batting average and its discontents

- Batting average is still widely cited as a measure of a hitter's quality today
- However, it quickly developed some critics!
- About 50 years after Chadwick popularized the batting average, another sports writer, F.C. Lane, began (repeatedly) noting its biggest flaw:

*Would a system that placed nickels, dimes, quarters, and 50-cent pieces on the same basis be much of a system whereby to compute a man's financial resources? And yet it is precisely such a loose, inaccurate system which obtains in baseball.*

– F.C. Lane (1916)

# Batting average and its discontents

- Batting average is still widely cited as a measure of a hitter's quality today
- However, it quickly developed some critics!
- About 50 years after Chadwick popularized the batting average, another sports writer, F.C. Lane, began (repeatedly) noting its biggest flaw:

*Would a system that placed nickels, dimes, quarters, and 50-cent pieces on the same basis be much of a system whereby to compute a man's financial resources? And yet it is precisely such a loose, inaccurate system which obtains in baseball.*

– F.C. Lane (1916)

- Arguing about sports stats has a long history!

# Alternative measures of hitting prowess

- Due to Lane's advocacy, statisticians developed the **slugging percentage** and **on-base percentage** as alternative hitting metrics

# Alternative measures of hitting prowess

- Due to Lane's advocacy, statisticians developed the [slugging percentage](#) and [on-base percentage](#) as alternative hitting metrics
- On-base percentage was invented in the 1950s by Dodgers executive Branch Rickey and his team statistician, Allan Roth



# Alternative measures of hitting prowess

- Due to Lane's advocacy, statisticians developed the [slugging percentage](#) and [on-base percentage](#) as alternative hitting metrics
- On-base percentage was invented in the 1950s by Dodgers executive Branch Rickey and his team statistician, Allan Roth
- After these “big three” measures of batter impact were adopted as official measures by MLB, two other measures followed:

# Alternative measures of hitting prowess

- Due to Lane's advocacy, statisticians developed the [slugging percentage](#) and [on-base percentage](#) as alternative hitting metrics
- On-base percentage was invented in the 1950s by Dodgers executive Branch Rickey and his team statistician, Allan Roth
- After these “big three” measures of batter impact were adopted as official measures by MLB, two other measures followed:
  1. [Batting run average](#), introduced in 1972 by two amateur baseball statisticians
  2. [On-base plus slugging](#) (OPS) was introduced in 1984 in a book by two other amateur baseball statisticians, one of whom became the Official Baseball Historian for Major League Baseball

# Alternative measures of hitting prowess

- Due to Lane's advocacy, statisticians developed the [slugging percentage](#) and [on-base percentage](#) as alternative hitting metrics
- On-base percentage was invented in the 1950s by Dodgers executive Branch Rickey and his team statistician, Allan Roth
- After these “big three” measures of batter impact were adopted as official measures by MLB, two other measures followed:
  1. [Batting run average](#), introduced in 1972 by two amateur baseball statisticians
  2. [On-base plus slugging](#) (OPS) was introduced in 1984 in a book by two other amateur baseball statisticians, one of whom became the Official Baseball Historian for Major League Baseball
- The development of new statistics was part of the goal of the [Society for American Baseball Research](#) in 1971, and for which the term [sabermetrics](#) is coined

# Moneyball

- Baseball had a rich analytics tradition, but analytics went mainstream with [Moneyball](#)

# Moneyball

- Baseball had a rich analytics tradition, but analytics went mainstream with [Moneyball](#)
- *Moneyball* tells the story of the early years of Billy Beane's tenure as GM of the Oakland Athletics (A's)

# Moneyball

- Baseball had a rich analytics tradition, but analytics went mainstream with *Moneyball*
- *Moneyball* tells the story of the early years of Billy Beane's tenure as GM of the Oakland Athletics (A's)
- In 1995, the As were bought by a relatively poor ownership group

# Moneyball

- Baseball had a rich analytics tradition, but analytics went mainstream with **Moneyball**
- *Moneyball* tells the story of the early years of Billy Beane's tenure as GM of the Oakland Athletics (A's)
- In 1995, the As were bought by a relatively poor ownership group
- Beane was hired in 1997 after several losing seasons

# Moneyball

- Baseball had a rich analytics tradition, but analytics went mainstream with [Moneyball](#)
- *Moneyball* tells the story of the early years of Billy Beane's tenure as GM of the Oakland Athletics (A's)
- In 1995, the As were bought by a relatively poor ownership group
- Beane was hired in 1997 after several losing seasons
- With his head statistician Paul DePodesta, Beane produced eight straight winning seasons beginning in 1999, despite having one of the lowest payrolls in MLB



# Moneyball

- Baseball had a rich analytics tradition, but analytics went mainstream with [Moneyball](#)
- *Moneyball* tells the story of the early years of Billy Beane's tenure as GM of the Oakland Athletics (A's)
- In 1995, the As were bought by a relatively poor ownership group
- Beane was hired in 1997 after several losing seasons
- With his head statistician Paul DePodesta, Beane produced eight straight winning seasons beginning in 1999, despite having one of the lowest payrolls in MLB
- The story of how they did this is detailed in Michael Lewis' book *Moneyball: The Art of Winning an Unfair Game* and a 2011 film starring Brad Pitt and Jonah Hill

# The Oakland A's 1998-2006

**Table 11.1** The Oakland A's Winning Percentage and Payroll: 1998–2006<sup>1</sup>

Year	Winning Percentage	Winning Percentage Rank	Payroll (in millions)	Payroll Rank
1998	0.457	21st	\$20.1	28th
1999	0.537	10th	\$24.2	26th
2000	0.565	6th	\$32.1	25th
2001	0.630	2nd	\$33.8	29th
2002	0.636	2nd	\$40.0	28th
2003	0.593	4th	\$50.3	23rd
2004	0.562	9th	\$59.4	16th
2005	0.543	9th	\$55.4	22nd
2006	0.574	5th	\$62.2	21st

1

- Because baseball has no payroll cap or floor, payroll discrepancies across teams can be huge
- Over the nine year stretch depicted, the A's paid *less than one third* of what the Yankees paid in total payroll

<sup>1</sup>From David Berri: *Sports Economics, First Edition*, 2018, Chapter 15

# How they did it (I)

- Evaluation of baseball prospects is based on tangible and intangible characteristics

# How they did it (I)

- Evaluation of baseball prospects is based on tangible and intangible characteristics
- Tangible characteristics are generally measurable while intangible characteristics are not

# How they did it (I)

- Evaluation of baseball prospects is based on tangible and intangible characteristics
- Tangible characteristics are generally measurable while intangible characteristics are not
- An experienced scout may be good at evaluating intangibles

# How they did it (I)

- Evaluation of baseball prospects is based on tangible and intangible characteristics
- Tangible characteristics are generally measurable while intangible characteristics are not
- An experienced scout may be good at evaluating intangibles
- For instance, in the movie, the scouts often express opinions and argue about players' innate "confidence", an intangible

# How they did it (I)

- Evaluation of baseball prospects is based on tangible and intangible characteristics
- Tangible characteristics are generally measurable while intangible characteristics are not
- An experienced scout may be good at evaluating intangibles
- For instance, in the movie, the scouts often express opinions and argue about players' innate "confidence", an intangible
- Beane and DePodesta *don't* claim to be better than average at sussing out intangibles, but instead focus a lot on a player's measurables: in particular their on base percentage

# How they did it (I)

- Evaluation of baseball prospects is based on tangible and intangible characteristics
- Tangible characteristics are generally measurable while intangible characteristics are not
- An experienced scout may be good at evaluating intangibles
- For instance, in the movie, the scouts often express opinions and argue about players' innate "confidence", an intangible
- Beane and DePodesta *don't* claim to be better than average at sussing out intangibles, but instead focus a lot on a player's measurables: in particular their on base percentage
- This leads to a fundamental conflict of methods between Beane/DePodesta and the scouts, which accounts for much of the drama in the movie



## How they did it (II)

- More specifically, Beane and DePodesta looked for evidence of **market inefficiencies** in the market for baseball players

## How they did it (II)

- More specifically, Beane and DePodesta looked for evidence of **market inefficiencies** in the market for baseball players
- Certain measurable characteristics, like the ability to take walks, are not sufficiently *compensated* in the player labour market

## How they did it (II)

- More specifically, Beane and DePodesta looked for evidence of **market inefficiencies** in the market for baseball players
- Certain measurable characteristics, like the ability to take walks, are not sufficiently *compensated* in the player labour market
- This is important, since it means that by hiring players who take a lot of walks, a low-budget team can buy more production and hence more wins per \$

## How they did it (II)

- More specifically, Beane and DePodesta looked for evidence of **market inefficiencies** in the market for baseball players
- Certain measurable characteristics, like the ability to take walks, are not sufficiently *compensated* in the player labour market
- This is important, since it means that by hiring players who take a lot of walks, a low-budget team can buy more production and hence more wins per \$
- It's hard for a small market team to succeed just on evaluating intangibles since the best scouts also likely work for bigger budget teams!

# Moneyball and Moreyball

- After Beane's success, other teams across major league sports started looking for similar market inefficiencies to exploit

# Moneyball and Moreyball

- After Beane's success, other teams across major league sports started looking for similar market inefficiencies to exploit
- In basketball, a version of Moneyball emerged around 2011-13 with what came to be called **Moreyball** after Houston Rockets (now Philadelphia) general manager Daryl Morey

# Moneyball and Moreyball

- After Beane's success, other teams across major league sports started looking for similar market inefficiencies to exploit
- In basketball, a version of Moneyball emerged around 2011-13 with what came to be called **Moreyball** after Houston Rockets (now Philadelphia) general manager Daryl Morey
- Analysts and executives like Morey noticed that, under the current NBA rules, teams were not taking enough threes, and three point shooting was undervalued in the player labour market

# Moneyball and Moreyball

- After Beane's success, other teams across major league sports started looking for similar market inefficiencies to exploit
- In basketball, a version of Moneyball emerged around 2011-13 with what came to be called **Moreyball** after Houston Rockets (now Philadelphia) general manager Daryl Morey
- Analysts and executives like Morey noticed that, under the current NBA rules, teams were not taking enough threes, and three point shooting was undervalued in the player labour market
- The resulting **three point revolution** in the NBA is an example of a Moneyball-style transformation, including the rise of a new class of stars like Steph Curry and James Harden



# What do we want a metric to do?

- Consider now the questions a sports data user asks himself when looking at a metric:

# What do we want a metric to do?

- Consider now the questions a sports data user asks himself when looking at a metric:

1. (General manager) Should I sign my free agent goalie to an extension based on this metric?
2. (Fan) Should I draft Bam Adebayo in the first round of my fantasy draft based on this metric?
3. (Gambler) Should I take the over or the under on the Vegas prediction of Mets' wins this season based on this metric?

# What do we want a metric to do?

- Consider now the questions a sports data user asks himself when looking at a metric:

1. (General manager) Should I sign my free agent goalie to an extension based on this metric?
2. (Fan) Should I draft Bam Adebayo in the first round of my fantasy draft based on this metric?
3. (Gambler) Should I take the over or the under on the Vegas prediction of Mets' wins this season based on this metric?

- In all cases, we need a metric to be two things:

# What do we want a metric to do?

- Consider now the questions a sports data user asks himself when looking at a metric:

1. (General manager) Should I sign my free agent goalie to an extension based on this metric?
2. (Fan) Should I draft Bam Adebayo in the first round of my fantasy draft based on this metric?
3. (Gambler) Should I take the over or the under on the Vegas prediction of Mets' wins this season based on this metric?

- In all cases, we need a metric to be two things:

1. **Relevant**: the metric correlates highly with wins

# What do we want a metric to do?

- Consider now the questions a sports data user asks himself when looking at a metric:

1. (General manager) Should I sign my free agent goalie to an extension based on this metric?
2. (Fan) Should I draft Bam Adebayo in the first round of my fantasy draft based on this metric?
3. (Gambler) Should I take the over or the under on the Vegas prediction of Mets' wins this season based on this metric?

- In all cases, we need a metric to be two things:

1. **Relevant**: the metric correlates highly with wins
2. **Stable over time**: the metric is predictive of the future

# Explaining game outcomes vs. explaining player skill

- We obviously want relevant metrics that can give of simple measures of a player's contribution to wins

# Explaining game outcomes vs. explaining player skill

- We obviously want relevant metrics that can give of simple measures of a player's contribution to wins
- But a player's contribution to wins can vary over time for reasons that are mainly due to luck (or role) rather than to innate skill

# Explaining game outcomes vs. explaining player skill

- We obviously want relevant metrics that can give of simple measures of a player's contribution to wins
- But a player's contribution to wins can vary over time for reasons that are mainly due to luck (or role) rather than to innate skill
  - This is especially true when a player's performance is highly dependent on his teammates



# Explaining game outcomes vs. explaining player skill

- We obviously want relevant metrics that can give of simple measures of a player's contribution to wins
- But a player's contribution to wins can vary over time for reasons that are mainly due to luck (or role) rather than to innate skill
  - This is especially true when a player's performance is highly dependent on his teammates
- It is important that a metric be mainly picking up the player's skill rather than luck, because in the latter case we could replace the player with any other player and have to the same *expected* on-field outcomes

# Explaining game outcomes vs. explaining player skill

- We obviously want relevant metrics that can give of simple measures of a player's contribution to wins
- But a player's contribution to wins can vary over time for reasons that are mainly due to luck (or role) rather than to innate skill
  - This is especially true when a player's performance is highly dependent on his teammates
- It is important that a metric be mainly picking up the player's skill rather than luck, because in the latter case we could replace the player with any other player and have to the same *expected* on-field outcomes
- One way to distinguish skill from luck is that **skill should be consistent over time while luck will change over time**

# Relevance vs. stability in sports analytics

- Problem: metrics that predict wins don't necessarily reflect skill

# Relevance vs. stability in sports analytics

- Problem: metrics that predict wins don't necessarily reflect skill
- This is especially true for athletes in other sports than baseball: so called **complex invasion sports** like hockey and basketball

# Relevance vs. stability in sports analytics

- Problem: metrics that predict wins don't necessarily reflect skill
- This is especially true for athletes in other sports than baseball: so called **complex invasion sports** like hockey and basketball
- For instance: an analysis by Berri and Schmidt (2010) found very low stability (i.e. year on year correlation) of goalie performance metrics in a sample of 361 NHL goalies

**Table 6.6** Consistency of NHL Goalies: 2000–01 to 2016–17

Statistic	$r$
Save percentage	0.23
Goals against average	0.31
Shots on goal, minute	0.60

- Goalie save percentage is obviously very important to a team's wins but it is very unstable for any goalie across seasons!

# How have sports statisticians and economists solved the relevance vs. stability problem?

🏀 We haven't!

# How have sports statisticians and economists solved the relevance vs. stability problem?

- We haven't!
- However, production and analysis of sports data continues to evolve, becoming more complex and predictive

# How have sports statisticians and economists solved the relevance vs. stability problem?

- We haven't!
- However, production and analysis of sports data continues to evolve, becoming more complex and predictive
- Some of these innovations include:



# How have sports statisticians and economists solved the relevance vs. stability problem?

- We haven't!
- However, production and analysis of sports data continues to evolve, becoming more complex and predictive
- Some of these innovations include:
  1. **Plus/minus data**: how well does the team do with an without a certain player on the field/court/ice

# How have sports statisticians and economists solved the relevance vs. stability problem?

- We haven't!
- However, production and analysis of sports data continues to evolve, becoming more complex and predictive
- Some of these innovations include:
  1. **Plus/minus data**: how well does the team do with an without a certain player on the field/court/ice
  2. **Tracking data**: where are players locating during offense and defense: how fast to they run? how many shots do they contest?

# How have sports statisticians and economists solved the relevance vs. stability problem?

- We haven't!
- However, production and analysis of sports data continues to evolve, becoming more complex and predictive
- Some of these innovations include:
  1. **Plus/minus data**: how well does the team do with an without a certain player on the field/court/ice
  2. **Tracking data**: where are players locating during offense and defense: how fast to they run? how many shots do they contest?
  3. **Machine learning and decay models**: how confident can we be that a breakout performance by an athlete is “real” based on prior information about similar players and performances?

# What about soccer?

- Extensive use of analytics has been relatively slow to come to the world of professional soccer, despite it being the largest revenue-generating sport in the world

# What about soccer?

- Extensive use of analytics has been relatively slow to come to the world of professional soccer, despite it being the largest revenue-generating sport in the world
- Simple analytics on the level of batting average are generally difficult to create in a sport with very little scoring

# What about soccer?

- Extensive use of analytics has been relatively slow to come to the world of professional soccer, despite it being the largest revenue-generating sport in the world
- Simple analytics on the level of batting average are generally difficult to create in a sport with very little scoring
- Instead, modern soccer analytics focus on **possession**: who controls the ball the most and what do they do with it?

# What about soccer?

- Extensive use of analytics has been relatively slow to come to the world of professional soccer, despite it being the largest revenue-generating sport in the world
- Simple analytics on the level of batting average are generally difficult to create in a sport with very little scoring
- Instead, modern soccer analytics focus on **possession**: who controls the ball the most and what do they do with it?
- Like a lot of modern stats in sports, understanding possessions requires extensive use of **tracking data**

# Modern possession-based soccer models

Key soccer possession models by publication year, with type of model and possession information

NAME	CREATOR	DEBUT	METHOD	WINDOW	OFF-BALL INFORMATION
Markov Chains	S. Rudd	2011	Markov chain	One possession	Defensive states tagged in event data
Possession-Based Model	N. Mackay	2016	Logistic regression and GAM	One possession	None
Expected Threat (xT)	K. Singh	2019	Markov-like	Next 5 actions (goal for)	None
Valuing Actions by Estimating Probabilities (VAEP)	KU Leuven DTAI	2019	Gradient-boosted trees	Next 10 actions (goal for or against)	Possession history proxies
Expected Possession Value (EPV)	J. Fernández et al.	2019	Multiple models	Next goal (for or against) or end of half	Full tracking data

2

<sup>2</sup>Graphic from FiveThirtyEight Sports, 2022



END

# Henry Chadwick



# Branch Rickey



# Moneyball

