# Finding, Monitoring, and Checking Claims Computationally Based on Structured Data

Brett Walenz, You (Will) Wu, Seokhyun (Alex) Song, Emre Sonmez, Eric Wu, Kevin Wu, Pankaj K. Agarwal, Jun Yang — Duke University

Naeemul Hassan, Afroza Sultana, Gensheng Zhang, Chengkai Li — University of Texas, Arlington

Cong Yu — Google, Inc.

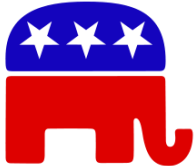**DUKE** COMPUTER SCIENCE  UNIVERSITY OF **TEXAS** ARLINGTON  Research at Google

# Claims based on data …



"Shaquille O'Neal had 40 points and 19 rebounds in the game against the Detroit Pistons on April 5, 1995. No one had a better performance in season 1994-95."
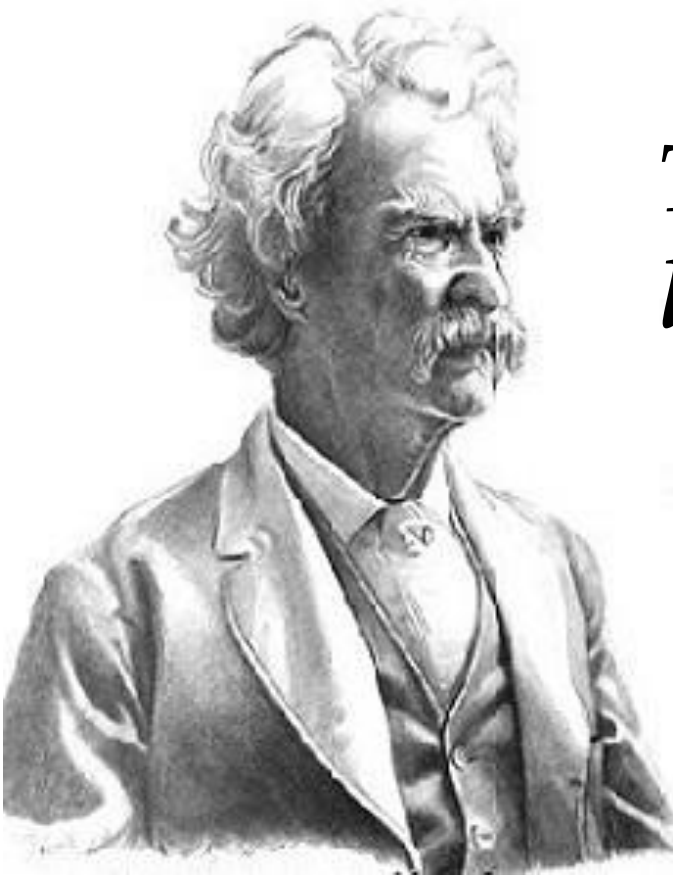


"During her six years in the Senate, Hagan has rubber-stamped the Obama agenda 95% of the time."



Jim Marshall, a Democratic incumbent from Georgia, voted with Nancy Pelosi "almost 90 percent of the time"



Jim Marshall "is a long way from Nancy Pelosi," as he "voted the same as Republican leaders 65 percent of the time"

*There are lies, damned lies, and statistics.*

– Mark Twain

How do we check these claims?

# Challenge: *vagueness*



"During her six years in the Senate, Hagan has rubber-stamped the Obama agenda 95% of the time."

Huh? "Obama agenda"? "95%"?

A lot of "hidden" information in here.

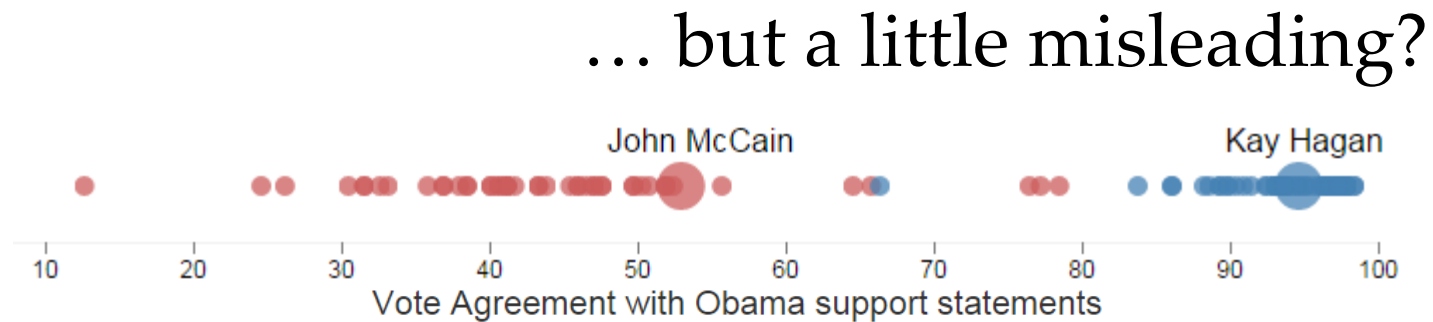"Obama agenda" : official statements made by President Obama about a bill OR nomination

"six years, 95% of the time": That sounds… bad? Is it? Does this mean all six years, or just lately?

# Challenge: *beyond correctness*

## Correct…

| NEW MEXICO | |
|---|---|
| Bingaman | 96 |
| Udall | 96 |
| **NEW YORK** | |
| Schumer | 99 |
| Gillibrand | 97 |
| **NORTH CAROLINA** | |
| **Burr** | 45 |
| Hagan | 96 |
| **NORTH DAKOTA** | |
| Conrad | 96 |
| **Hoeven** | 60 |
| **OHIO** | |
| Brown | 96 |
| **Portman** | 49 |

… but a little misleading?



John McCain          Kay Hagan

Vote Agreement with Obama support statements

Source: Congressional Quarterly

# Challenge: *examine counter arguments*



"During her six years in the Senate, Hagan has rubber-stamped the Obama agenda 95% of the time."
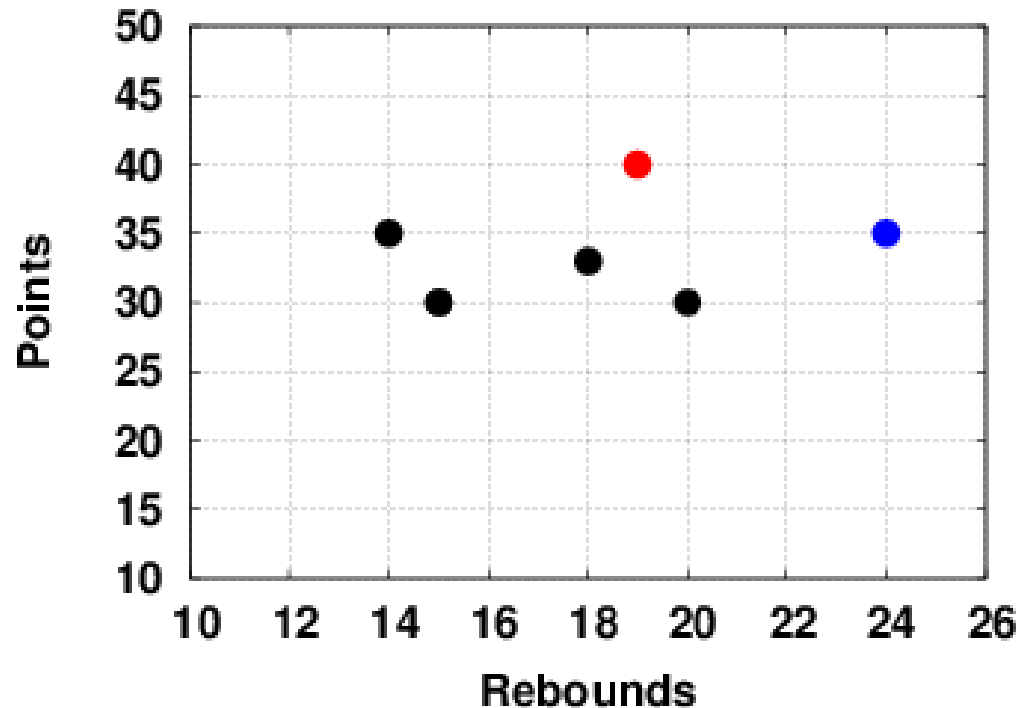
**Counter-argument**



"During the years 2012-2013, Democrats on average voted 94% of the time in line with Obama's public position. Kay Hagan votes within 1% of the average Democrat on Obama's position."

# Challenge: *generating claims*

"Shaquille O'Neal had 40 points and 19 rebounds in the game against the Detroit Pistons on April 5, 1995. No one had a better performance in season 1994-95."

# Goal

- Fact-checking is growing by leaps and bounds, can be aided by analytic process
- How much can we automate this process?
  - Can we quantify quality beyond correctness?
  - Can we formulate
    - reverse-engineering of vague claims
    - finding counterarguments
    - generating/monitoring claims

    as computational problems?
  - Can we do so in a general way, for many claims in many domains?

To check a claim, **tweak** the way it manipulates data and see if we get different conclusions.

Obama agenda

Bills
Nominations
General Votes

"During her six years in the Senate, Hagan has rubber-stamped the Obama agenda 95% of the time."

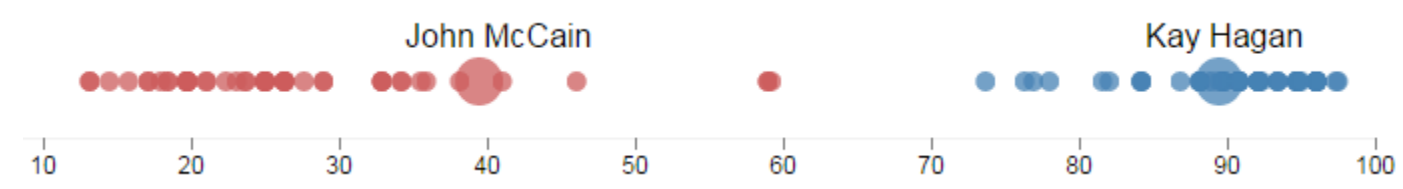2009-2014

Date Ranges:
2009-2010,
2010-2011,
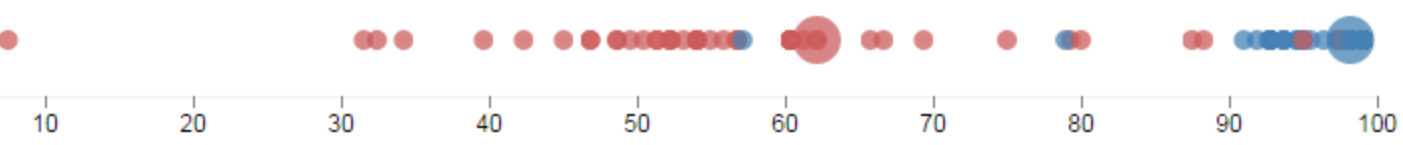2012-2014, …

Hagan

Democrats
Republicans
Individuals
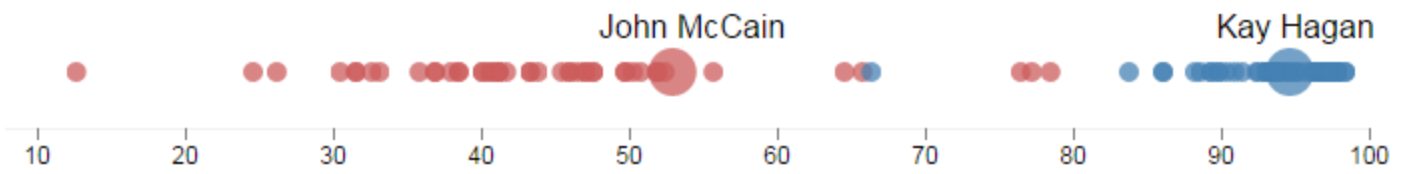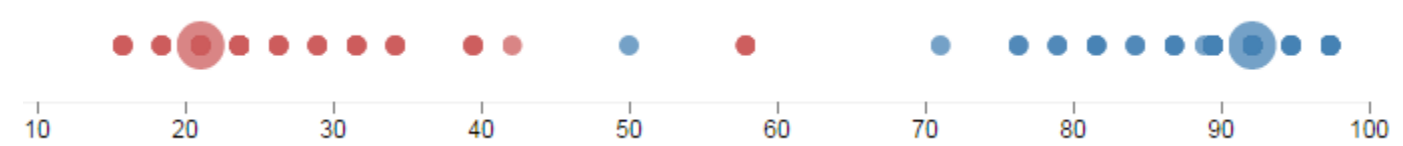
Vote agreement with a public Obama position
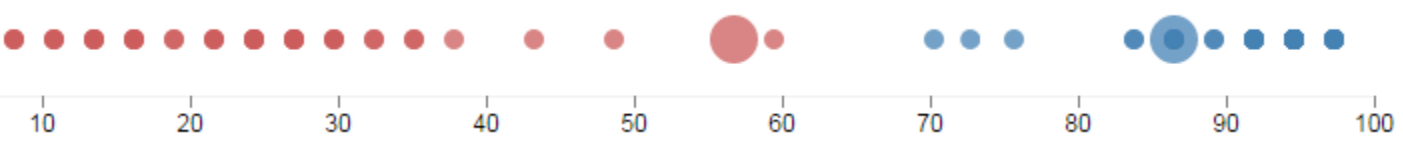
# To generate a claim, **tweak** the way it manipulates data and see *how* it compares to others

*D*         *M*

| Id | Player | Team | Season | Points | Assists | Blocks |
|----|--------|------|--------|--------|---------|--------|
| $t_1$ | Lamar Odom | LA Clippers | 2003 | 30 | 11 | 12 |
| $t_2$ | Eddie House | Miami Heat | 2003 | 29 | 7 | 8 |
| $t_3$ | Eddie House | Miami Heat | 2003 | 28 | 6 | 9 |
| $t_4$ | Lamar Odom | Miami Heat | 2004 | 32 | 9 | 13 |
| $t_5$ | Lamar Odom | Miami Heat | 2004 | 28 | 11 | 6 |
| $t_6$ | Lamar Odom | Miami Heat | 2004 | 29 | 9 | 7 |
| $t_7$ | Eddie House | LA Clippers | 2004 | 30 | 11 | 10 |
| $t_8$ | Lamar Odom | Miami Heat | 2004 | 30 | 11 | 11 |

| Conditions | Combinations |
|------------|--------------|
| Season = 2004 | Assists, Blocks |
| Player = Lamar Odom Season = 2004 | Points, Assists |
| … | … |

Find conditions over *D* and combinations of *M* such that
- $t_8$ is in the skyline
- $t_8$ generates a prominent streak

- Lamar Odom scored 11 assists and 11 blocks. No one made a better performance in season 2004

- Lamar Odom had at least 28 points and 9 or more assists for 4 consecutive games; his the longest such streak in 2004

# Parameterized Queries

- A claim, such as the Kay Hagan example, is a *template* with a set of existing *parameters* (*ex. Obama agenda, six years, Kay Hagan)*

- iCheck/uClaim works on *query templates*, which tell us how to *get* the data and *perturb* the parameters

- In addition, we need to understand how to compare and contrast results and parameters

# Parameterized Queries II

- Relative claim strength – a function to determine how to compare results (*ex. lower is better in the Kay Hagan example)*

- Not all parameter perturbations are sensible (*ex. dates before Obama took office).* Need a *parameter sensibility* function

# uClaim

# uClaim

[April 24, 1994] David Robinson had 71 points and 14 rebounds in the San Antonio Spurs' victory against the Los Angeles Clippers. No one before had a better performance in NBA history.

[April 20, 1994] Shaquille O'neal had 53 points and 18 rebounds in the Orlando Magic's win over the Minnesota Timberwolves. No one before had a better performance in NBA history.

[February 16, 1993] Shaquille O'neal had 46 points and 21 rebounds in the Orlando Magic's defeat against the Detroit Pistons. No one before had a better performance in NBA history.
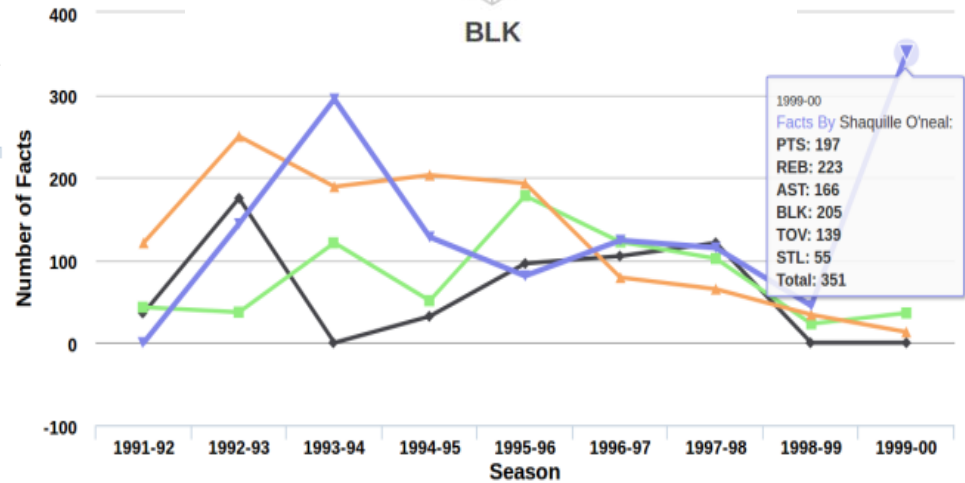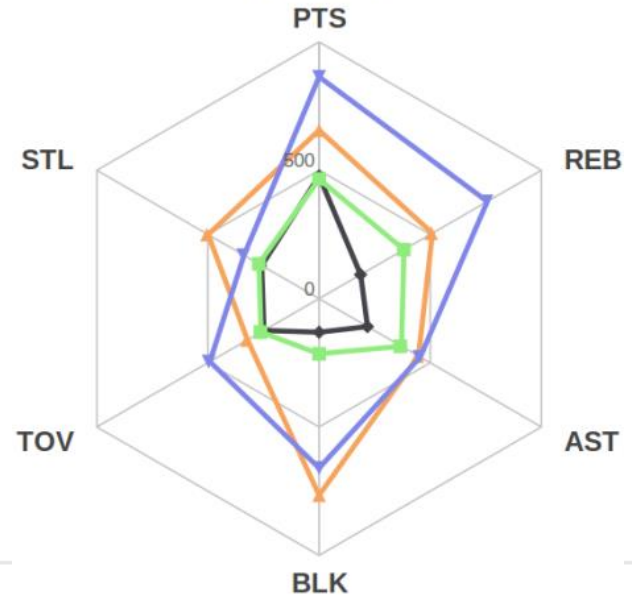
[February 27, 1992] David Robinson had 37 points and 24 rebounds in the San Antonio Spurs' victory against the Golden State Warriors. No one before had a better performance in NBA history.

## Compare Similar Stories
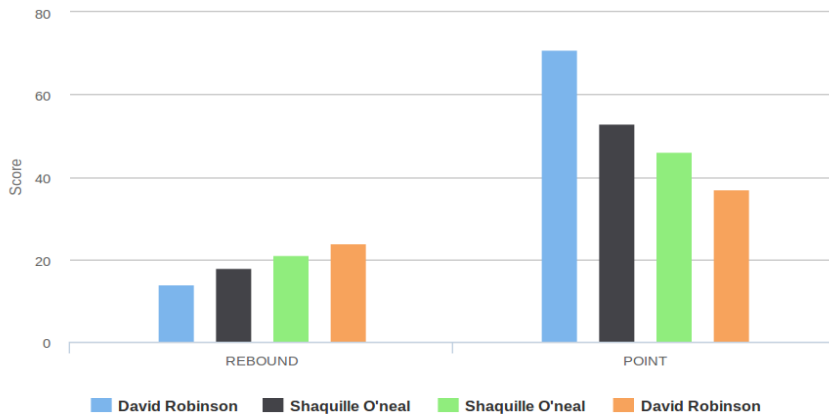
Similar Stories

## Number of Facts

Comparison of Players

# iCheck

✔**Kay Hagan is overly partisan.**

Republicans suggest her achievements are thin soup. "The only thing Kay Hagan has accomplished in Washington is becoming an automatic 'yes' vote for whatever new tax or regulation President Obama wants," North Carolina GOP Chairman Claude Pope said.

Original claim made by: **1**

## Supporting Arguments

| 7 | ↑ ↓ | Frank R. Lautenberg(D) and Kay Hagan(D) agreed on 90.02% of the votes they cast between 2011-01-05 (start of session 2011) and 2013-01-01 (end of session 2012) |
|---|---|---|
| 3 | ↑ ↓ | Thomas Harkin(D) and Kay Hagan(D) agreed on 92.6% of the votes they cast between 2013-01-03 (start of session 2013) and 2014-05-09 (end of session 2014) |

## Counter Arguments

## Generated Counter Arguments

| + | Bernard Sanders(Independent) and Kay Hagan(D) agreed on 85.77% of the votes they cast between 2011-01-05 (start of session 2011) and 2013-01-01 (end of session 2012) |
|---|---|
| + | Kay Hagan(D) voted 89.83% of the time with the Democrat party majority vote between 2008-01-02 and 2012-01-02. |

# Finding counter arguments



Vote Correlation

Vote agreement between Nancy Pelosi and Paul Ryan over varying time periods of 2009.
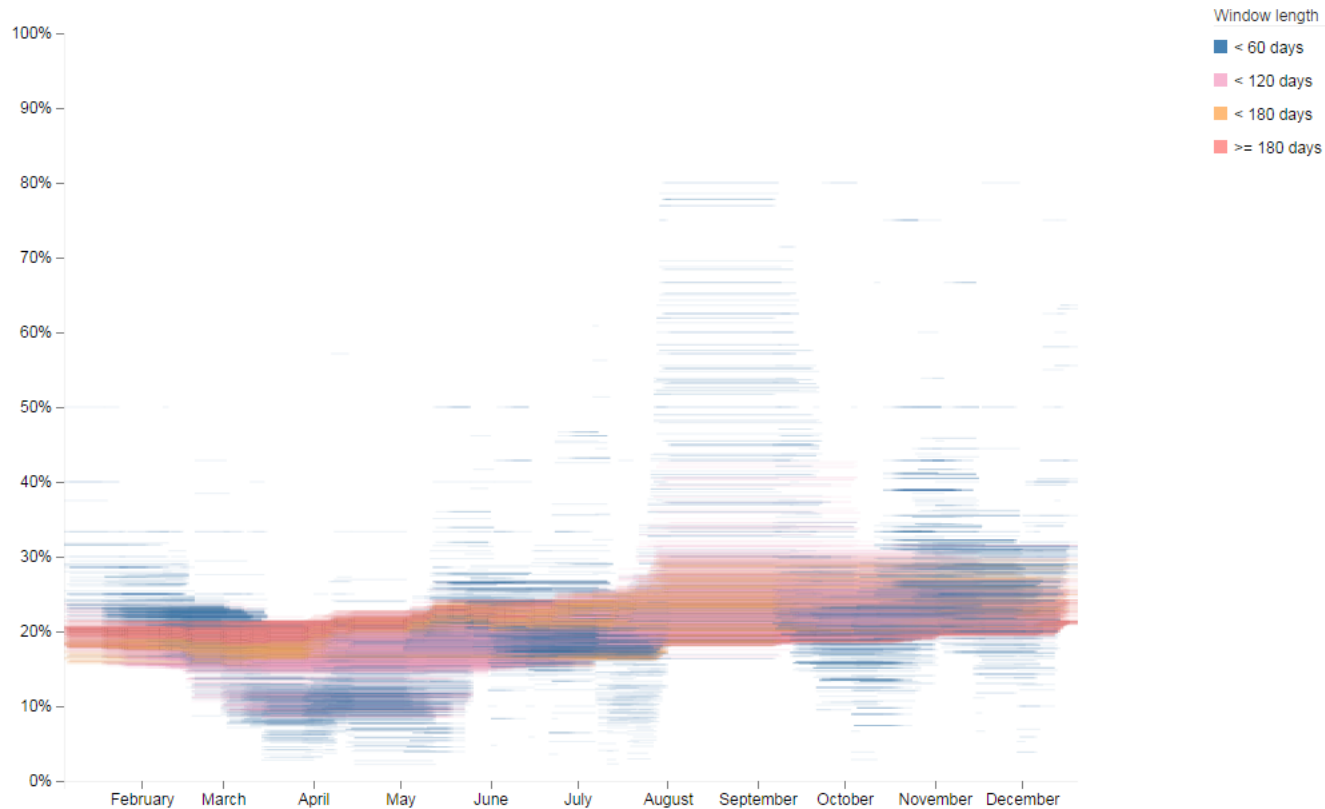
# iCheck



**Alex Gordon**

Views: 4

Bats: L, Throws: R | Kansas City Royals

| | |
|---|---|
| **Birth Date** | 1984-02-10 |
| **Birthplace** | Lincoln, NE, USA |
| **MLB Debut** | 2007-04-02 |
| **Final Game** | Currently Active |
| **Played under** | Buddy Bell, Trey Hillman, show all |
| **HT/WT** | 6'2", 220 lbs. |

**Awards**

| 2013 | Gold Glove |
|------|-----------|
| 2012 | Gold Glove |
| 2011 | Gold Glove |

## One-Of-The-Few Claims

👍 Bump!

🐦 Tweet!

In 2011, Alex Gordon had 185 in hits (batting), 23 in homeruns, 9 in GIDP; only 20 other players have ever beaten this record;

Views: 2 | Bumps: 0 | Tweets: 0

›

## Statistical Percentile Visualization

## Player Statistics

# iCheck

## Buddy Bell

Manager | Kansas City Royals

| | |
|---|---|
| **Birth Date** | 1951-08-27 |
| **Birthplace** | Pittsburgh, PA, USA |
| **HT/WT** | 6'1", 180 lbs. |
| **Managed** | Detroit Tigers, Colorado Rockies, show all |

### gate Claims

Bell managed the Detroit Tigers from 1996 to 1998, the Detroit Tigers's record dropped 1.5432% (from 1995 to 1998)
nter-arguments

Bell managed the Colorado Rockies from 2000 to 2002, the Colorado Rockies's record improved 0.6173% (from 1999 to 2002

Bell managed the Kansas City Royals from 2005 to 2007, the Kansas City Royals's record improved 6.7901% (from 2004 to 20
nter-arguments

Brian McRae, who ranked in the bottom 14.66% of all MLB players in strikeouts (105) in 1993, INCREASED Kansas City Royals's wins 7.41%! (153 games batted)
Mark Teahen, who ranked in the bottom 9.15% of all MLB players in strikeouts (258) from 2007 to 2008, INCREASED Kansas City Royals's wins 8.02%! (293 games
Alex Gordon, who ranked in the bottom 9.86% of all MLB players in strikeouts (257) from 2007 to 2008, INCREASED Kansas City Royals's wins 8.02%! (285 games
Bo Jackson, who ranked in the bottom 1.11% of all MLB players in strikeouts (476) from 1987 to 1989, INCREASED Kansas City Royals's wins 9.88%! (375 games batt
Danny Tartabull, who ranked in the bottom 7.78% of all MLB players in strikeouts (378) from 1987 to 1989, INCREASED Kansas City Royals's wins 9.88%! (437 game

# Thank you!

# Questions?

Brett Walenz, You (Will) Wu, Seokhyun (Alex) Song, Emre Sonmez, Eric Wu, Kevin Wu, Pankaj K. Agarwal, Jun Yang — Duke University

Naeemul Hassan, Afroza Sultana, Gensheng Zhang, Chengkai Li — University of Texas, Arlington

Cong Yu — Google, Inc.