



UNIVERSITY  
of VIRGINIA

CURRY SCHOOL *and*  
BATTEN SCHOOL  
EdPolicyWorks

# Introduction to Data Visualization

A Review of Principles, Techniques, Tools,  
and Resources

Brian Heseung Kim  
University of Virginia

June 25, 2019

EdPolicyWorks: The Center on Education Policy and Workforce Competitiveness

# Lesson Outline

---

- Setting the Stage
  - What is Data Visualization?
- Frameworks and Principles
  - A (Too) Brief Primer
- Getting Started
  - Taking a Test Drive
- Next Steps
  - Tools to Learn
  - Additional Resources

# Setting the Stage

# First Things First

---

- Please go to the following links to download these slides and a sample dataset.
  - <http://bit.ly/surpdatavizslides>
- Open the slides to access example links throughout the lesson

# What is Data Visualization?

---

Communicating

Information

Visually

# What is Data Visualization?

---

Communicating

~~Displaying~~

Information

Visually

# What is Data Visualization?

---

Communicating

~~Showing~~

Information

Visually

# What is Data Visualization?

---

Communicating

~~Presenting~~

Information

Visually

# What is Data Visualization?

---

Communicating

Information

Visually

# An Illustrative Example

---

- Please go to the following link:
  - <https://chriswhong.github.io/nyctaxi/>
- As you view, think about the following questions:
  - What insight or information is being communicated?  
In other words, what are you learning from this?
  - What questions do you have after seeing this visualization?
  - What strengths and weaknesses do you notice about this style of visualization?

# An Illustrative Example

## Data fields

- **id** - a unique identifier for each trip
- **vendor\_id** - a code indicating the provider associated with the trip record
- **pickup\_datetime** - date and time when the meter was engaged
- **dropoff\_datetime** - date and time when the meter was disengaged
- **passenger\_count** - the number of passengers in the vehicle (driver entered value)
- **pickup\_longitude** - the longitude where the meter was engaged
- **pickup\_latitude** - the latitude where the meter was engaged
- **dropoff\_longitude** - the longitude where the meter was disengaged
- **dropoff\_latitude** - the latitude where the meter was disengaged
- **store\_and\_fwd\_flag** - This flag indicates whether the trip record was held in vehicle memory before sending to the vendor because the vehicle did not have a connection to the server - Y=store and forward; N=not a store and forward trip
- **trip\_duration** - duration of the trip in seconds
- **fare\_amount** - `float` dollar amount of the cost of the taxi ride.

# An Illustrative Example

id	vendor_id	pickup_datetime	dropoff_datetime	passenger_count	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	store_a
id2875421	2	2016-03-14 17:24:55	2016-03-14 17:32:30	1	-73.98215	40.76794	-73.96463	40.76560	N
id2377394	1	2016-06-12 00:43:35	2016-06-12 00:54:38	1	-73.98042	40.73856	-73.99948	40.73115	N
id3858529	2	2016-01-19 11:35:24	2016-01-19 12:10:48	1	-73.97903	40.76394	-74.00533	40.71009	N
id3504673	2	2016-04-06 19:32:31	2016-04-06 19:39:40	1	-74.01004	40.71997	-74.01227	40.70672	N
id2181028	2	2016-03-26 13:30:55	2016-03-26 13:38:10	1	-73.97305	40.79321	-73.97292	40.78252	N
id0801584	2	2016-01-30 22:01:40	2016-01-30 22:09:03	6	-73.98286	40.74220	-73.99208	40.74918	N
id1813257	1	2016-06-17 22:34:59	2016-06-17 22:40:40	4	-73.96902	40.75784	-73.95741	40.76590	N
id1324603	2	2016-05-21 07:54:58	2016-05-21 08:20:49	1	-73.96928	40.79778	-73.92247	40.76056	N
id1301050	1	2016-05-27 23:12:23	2016-05-27 23:16:38	1	-73.99948	40.73840	-73.98579	40.73281	N
id0012891	2	2016-03-10 21:45:01	2016-03-10 22:05:26	1	-73.98105	40.74434	-73.97300	40.78999	N
id1436371	2	2016-05-10 22:08:41	2016-05-10 22:29:55	1	-73.98265	40.76384	-74.00223	40.73299	N
id1299289	2	2016-05-15 11:16:11	2016-05-15 11:34:59	4	-73.99153	40.74944	-73.95654	40.77063	N
id1187965	2	2016-02-19 09:52:46	2016-02-19 10:11:20	2	-73.96298	40.75668	-73.98441	40.76072	N
id0799785	2	2016-06-01 20:58:29	2016-06-01 21:02:49	1	-73.95631	40.76794	-73.96611	40.76300	N
id2900608	2	2016-05-27 00:43:36	2016-05-27 01:07:10	1	-73.99220	40.72723	-73.97466	40.78307	N
id3319787	1	2016-05-16 15:29:02	2016-05-16 15:32:33	1	-73.95551	40.76859	-73.94876	40.77155	N
id3379579	2	2016-04-11 17:29:50	2016-04-11 18:08:26	1	-73.99117	40.75556	-73.99929	40.72535	N
id1154431	1	2016-04-14 08:48:26	2016-04-14 09:00:37	1	-73.99426	40.74580	-73.99966	40.72334	N
id3552682	1	2016-06-27 09:55:13	2016-06-27 10:17:10	1	-74.00398	40.71301	-73.97919	40.74992	N

# Frameworks and Principles

# Strong Visualizations *should*...

---

- ✓ Have a focused purpose or intention
- ✓ Reveal insight, learnings, or ideas to the audience
- ✓ Be irreplaceable with words alone
- ✓ Be responsive to the intended audience
- ✓ Represent the data accurately, truthfully, and objectively

# Strong Visualizations *should*...

---

- ✓ Have a focused purpose or intention
- ✓ Reveal insight, learnings, or ideas to the audience
- ✓ Be irreplaceable with words alone
- ✓ Be responsive to the intended audience
- ✓ Represent the data accurately, truthfully, and objectively

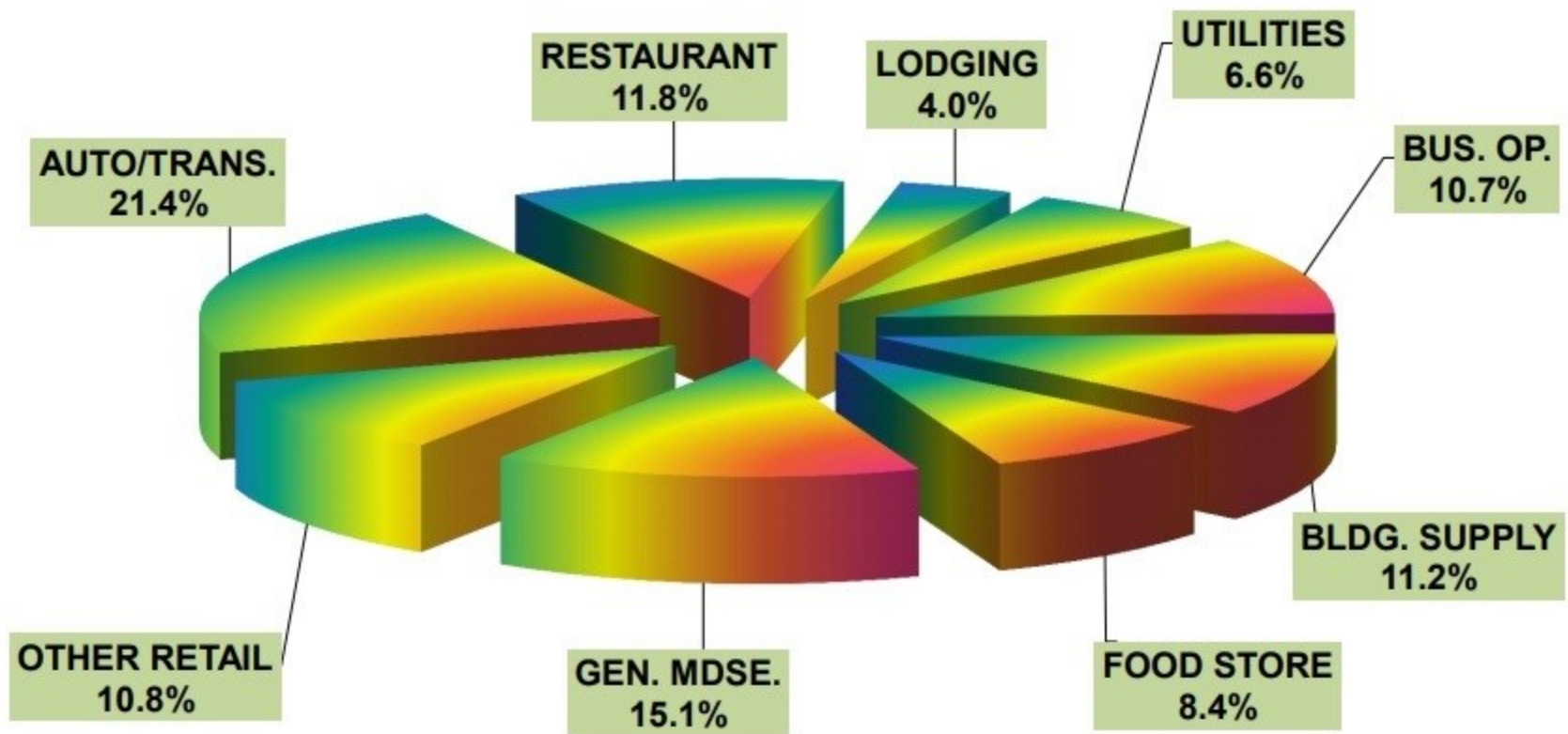
# The Largest Vocabulary in Hip-Hop

# Strong Visualizations *should*...

---

- ✓ Have a focused purpose or intention
- ✓ Reveal insight, learnings, or ideas to the audience
- ✓ Be irreplaceable with words alone
- ✓ Be responsive to the intended audience
- ✓ Represent the data accurately, truthfully, and objectively

## ***2015 Maine Taxable Sales by Sector***



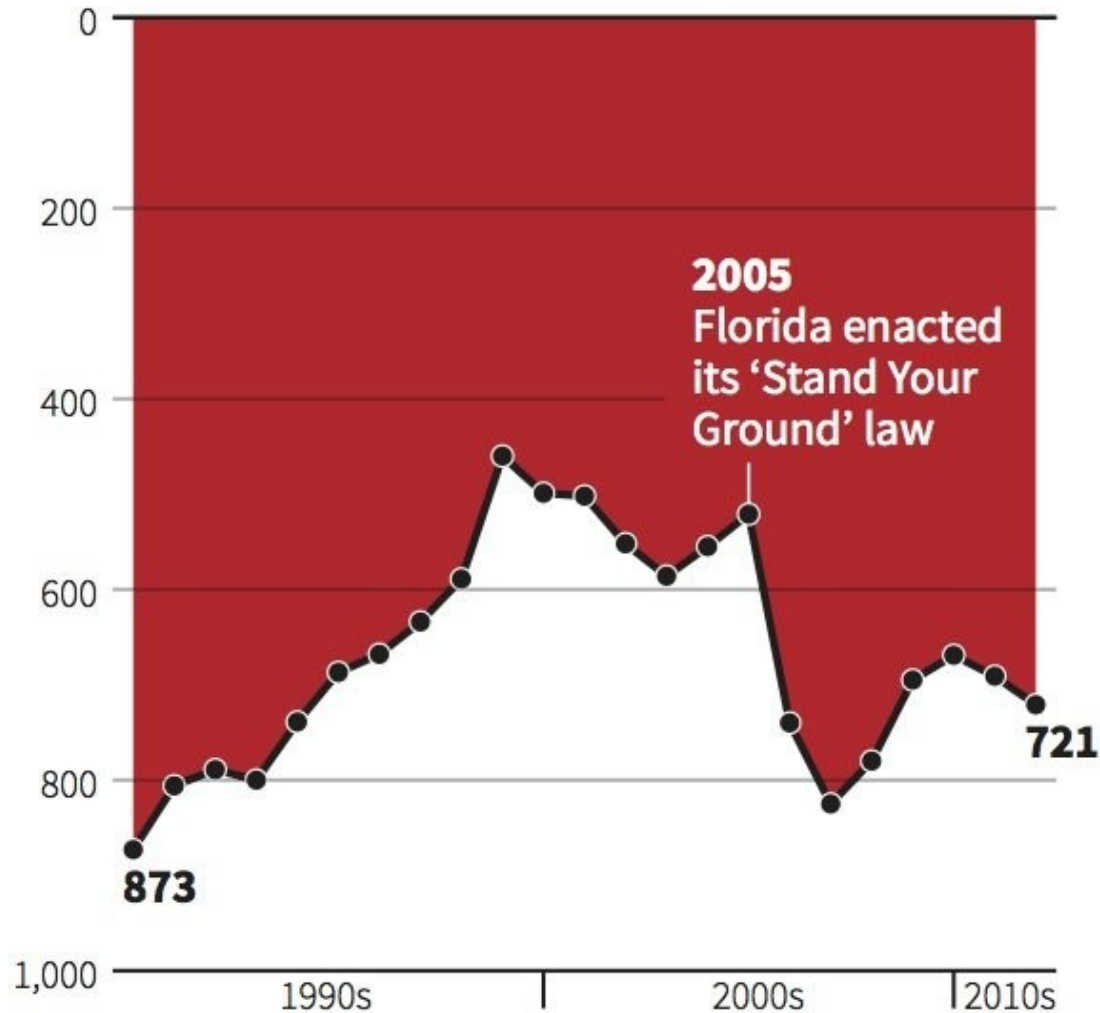
<https://legislature.maine.gov/uploads/originals/2016-december-report-final-2.pdf>

# Virginia School Quality Profiles

## A Proposed Revision

# Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

Source: [Heap.io](http://Heap.io)

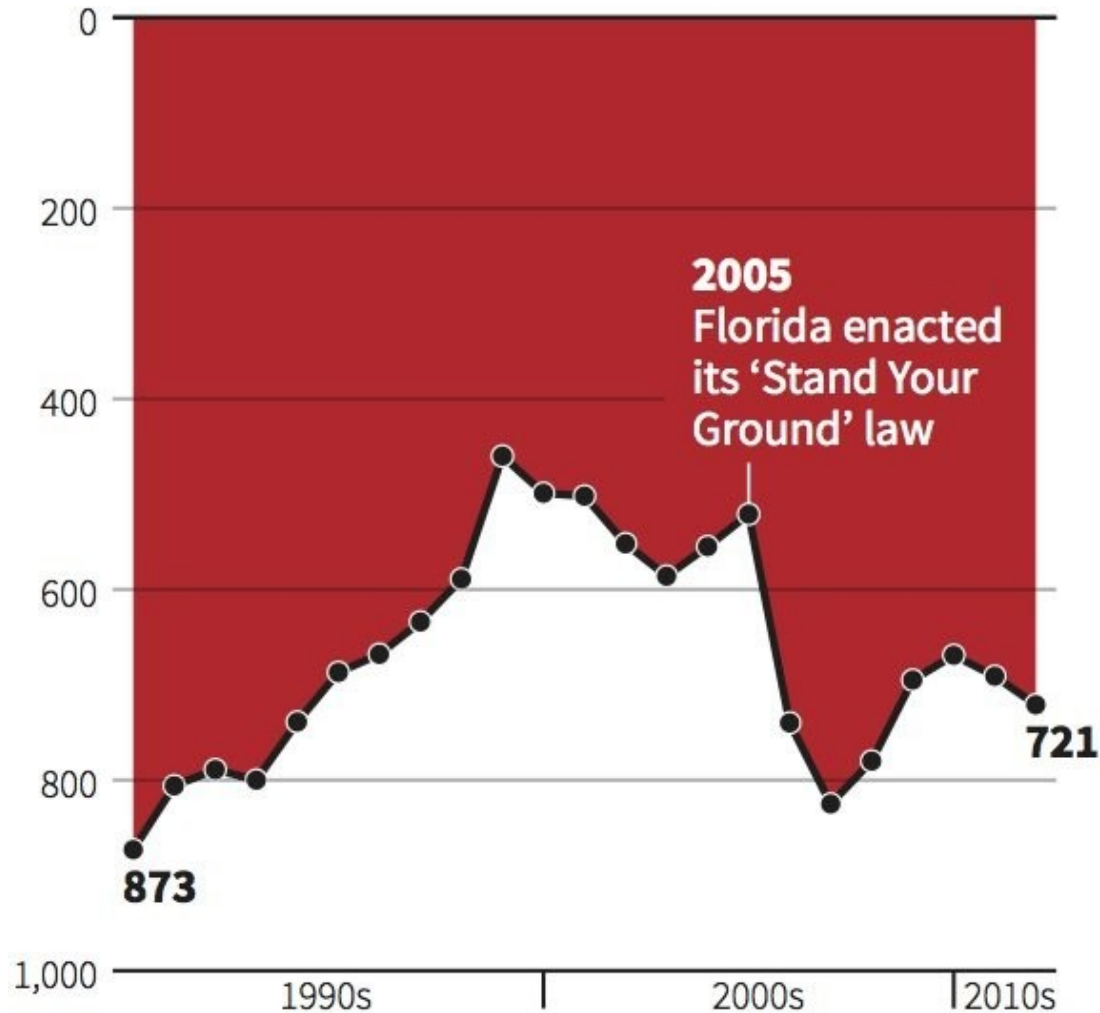
# Strong Visualizations *should*...

---

- ✓ Have a focused purpose or intention
- ✓ Reveal insight, learnings, or ideas to the audience
- ✓ Be irreplaceable with words alone
- ✓ Be responsive to the intended audience
- ✓ Represent the data accurately, truthfully, and objectively

# Gun deaths in Florida

Number of murders committed using firearms

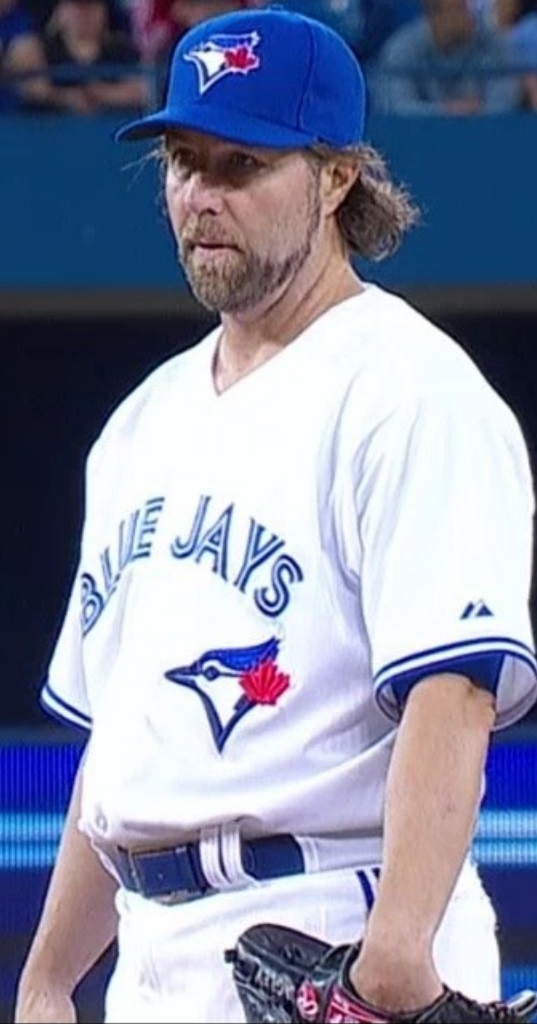
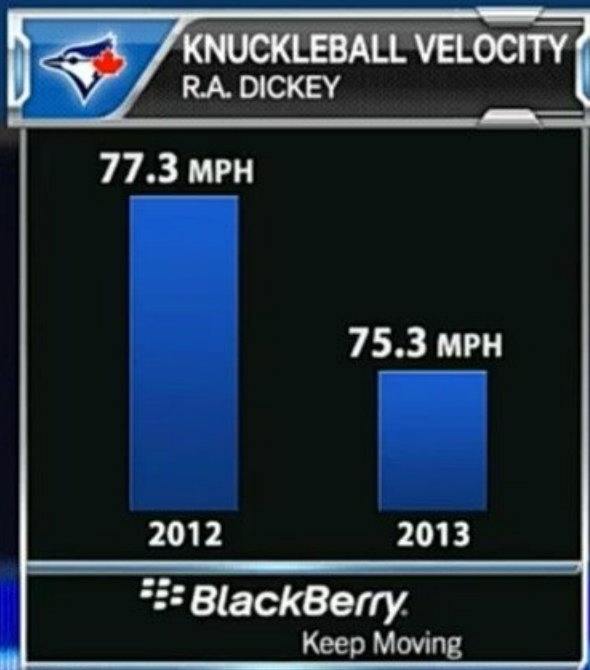


Source: Florida Department of Law Enforcement

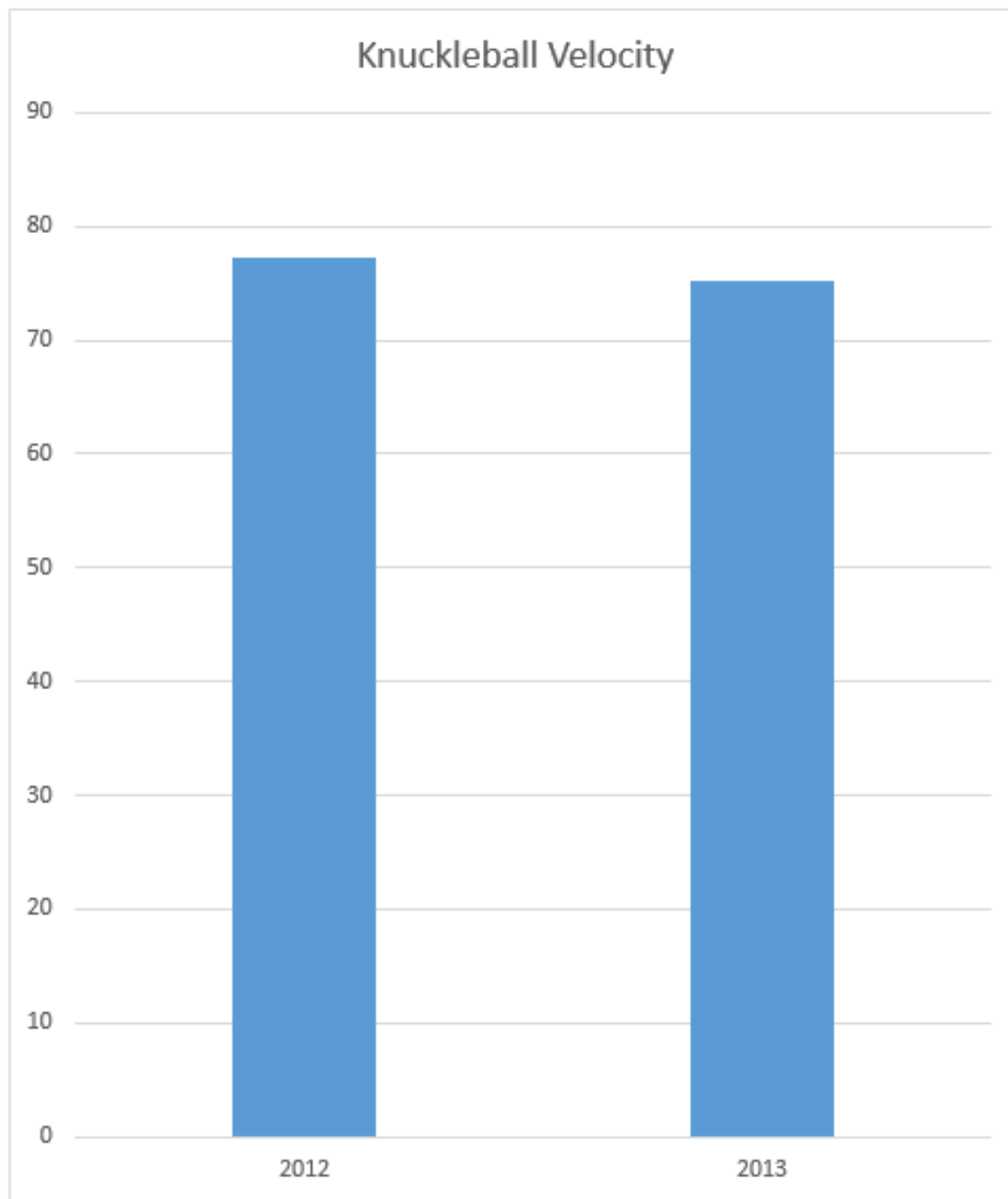
C. Chan 16/02/2014

REUTERS

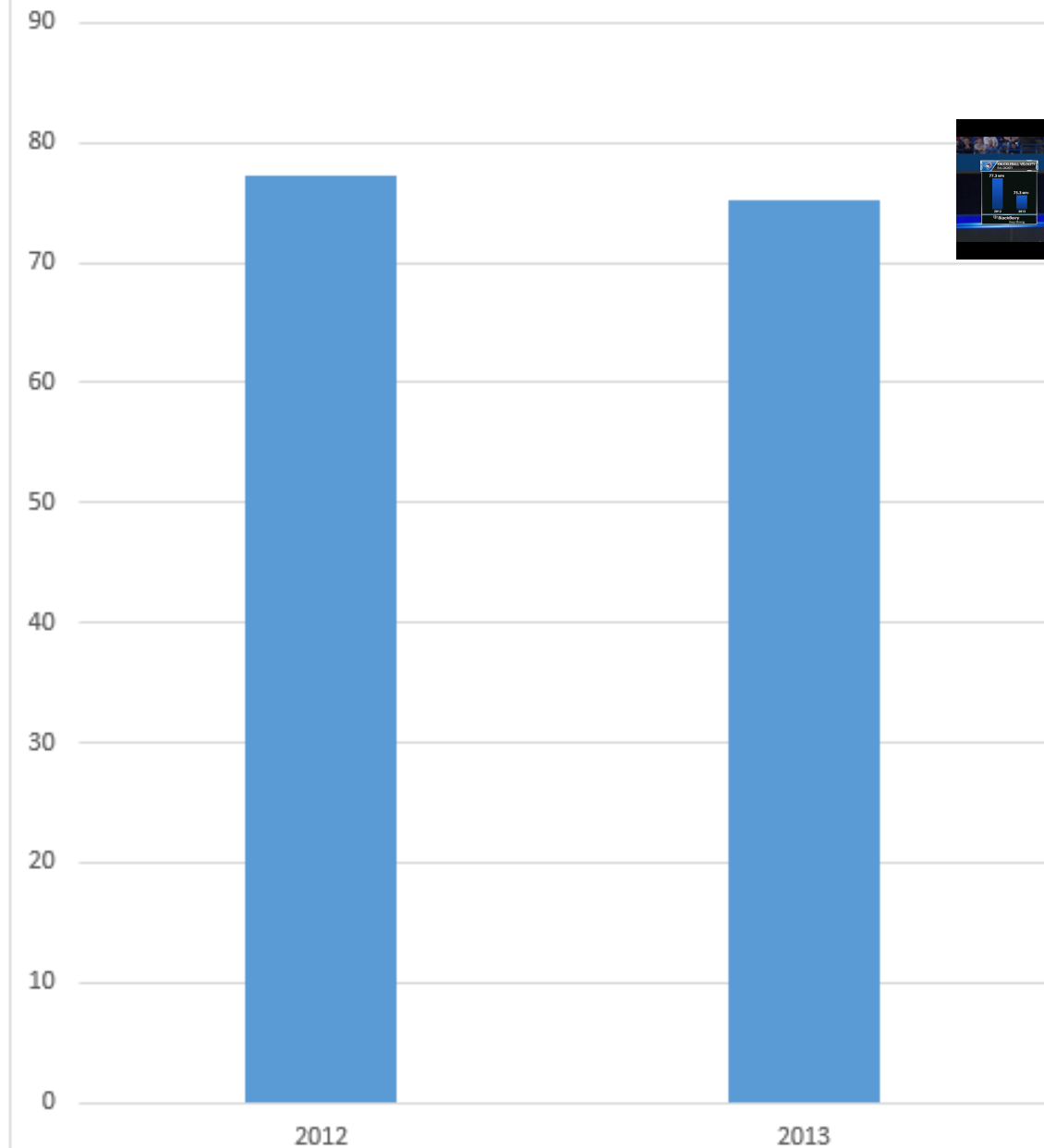
Source: [Heap.io](https://www.heap.io)



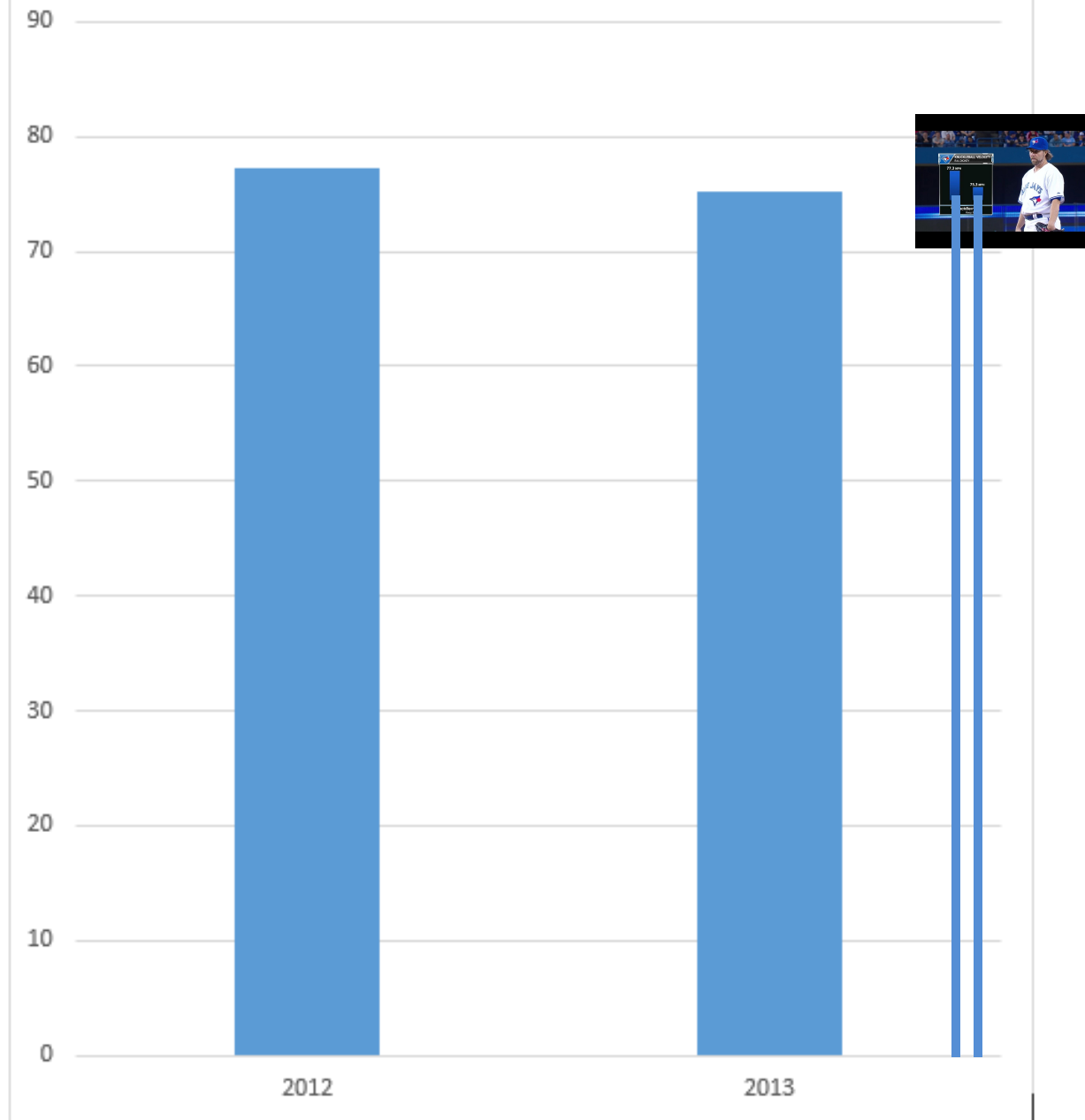
Source: [Heap.io](http://Heap.io)



## Knuckleball Velocity



## Knuckleball Velocity



More Great Examples Here

# Getting Started

# Taking a Test Drive

---

- Please go to the following link:
  - <http://app.rawgraphs.io/>
- As you experiment, think about the following questions:
  - What interesting relationships might exist in these data?
  - How can I best explore those relationships?
  - What further questions should I explore?

# Next Steps

# Tools to Learn (a.k.a. Pick your Poison)

---

- Data Prep:
  - Python, R, Tableau Prep are all free
- Less flexible viz, but more approachable:
  - Excel / Google Sheets
  - Rawgraphs.io
  - Tableau (free for students!)
- Very flexible viz, but steep learning curve:
  - R: ggplot2, plotly, leaflet
  - Python: Matplotlib and Seaborn
  - Javascript: D3 (“Data-Driven Documents”)

# Additional Resources

---

- For Inspiration:

- <https://flowingdata.com/>
- <https://observablehq.com/collection/@observablehq/visualization>
- <https://datavizcatalogue.com/index.html>
- <https://www.d3-graph-gallery.com/>
- <https://www.tableau.com/about/blog>
- <https://github.com/d3/d3/wiki/Gallery>
- <https://bl.ocks.org/>
- <http://www.visualisingdata.com/>
- <https://junkcharts.typepad.com/>
- <http://www.storytellingwithdata.com/>

- For Consumption:

- <https://pudding.cool/>
- <https://www.nytimes.com/section/upshot>
- <https://fivethirtyeight.com/>
- <https://www.economist.com/graphic-detail/>
- <https://www.census.gov/dataviz/>

# Any questions?

---

- Setting the Stage
  - What is Data Visualization?
- Frameworks and Principles
  - A (Too) Brief Primer
- Getting Started
  - Taking a Test Drive
- Next Steps
  - Tools to Learn
  - Additional Resources



UNIVERSITY  
of VIRGINIA

CURRY SCHOOL *and*  
BATTEN SCHOOL  
EdPolicyWorks

# Thank you!

Brian Heseung Kim



@brhkim



brian.kim  
@virginia.edu



brhkim.com



@brhkim

For the most up-to-date information from EdPolicyWorks,  
please contact us at:  
EdPolicyWorks@virgina.edu