

DATA SCIENCE TOOLBOX MEETING – WEEK 3

Brooke
Anderson

**ARE “HER”-ICANES
WORSE?**

SAVING LIVES DURING HURRICANES

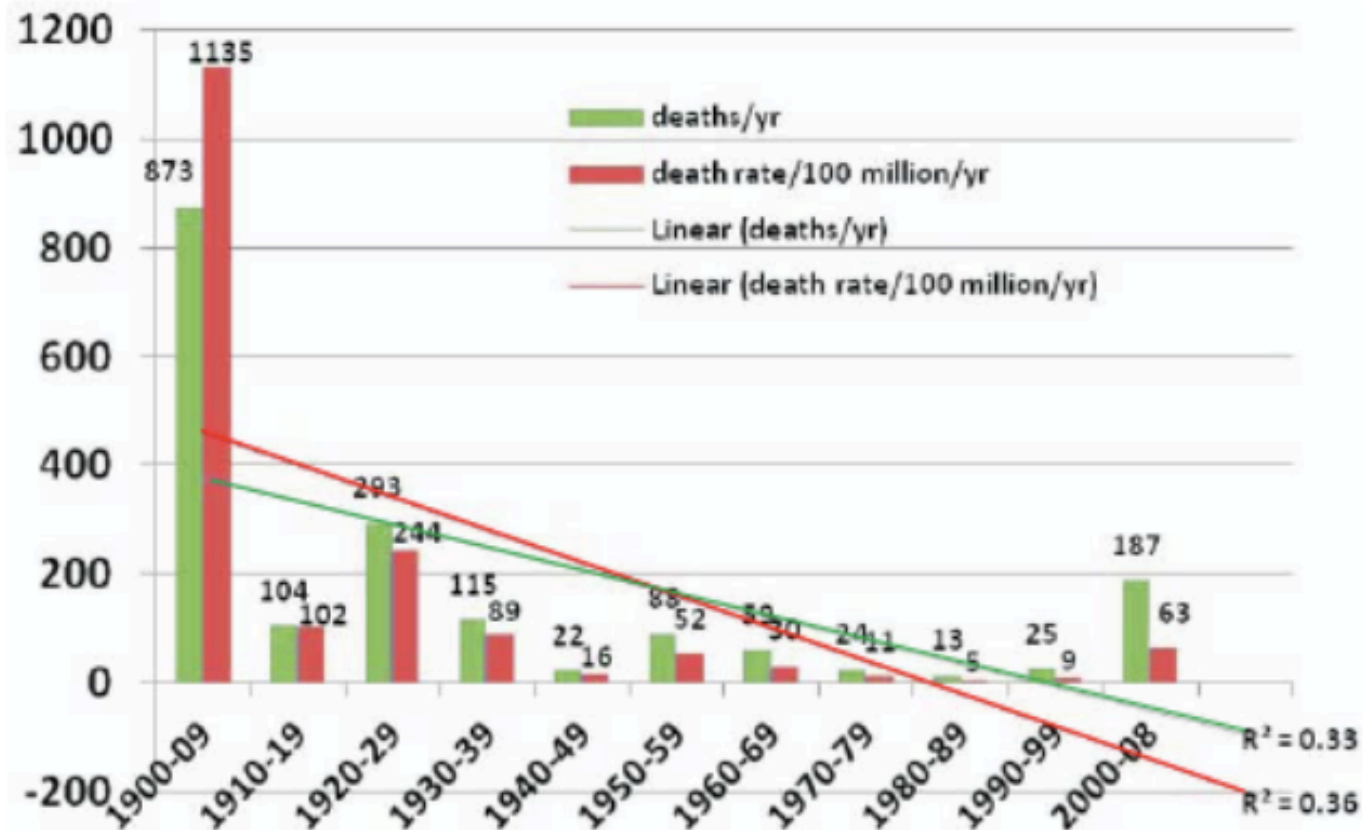


Figure 10. U.S. Deaths and Death Rates due to Hurricanes, 1900–2006. ^{31,21,36-38}

Source: Goklany, 2009 <http://www.jpands.org/vol14no4/goklany.pdf>



TAKING EVACUATIONS SERIOUSLY

Sources:

[http://
www.nbcnews.co
m/id/9439537/
ns/us_news-
katrina_the_long
_road_back/t/
staying-better-
write-your-ss-
your-body/
#.U6G5oxbePzl](http://www.nbcnews.com/id/9439537/ns/us_news-katrina_the_long_road_back/t/staying-better-write-your-ss-your-body/#.U6G5oxbePzl),

Wikipedia
Commons

Staying? Better write your SS# on your body...

From one mayor, a sobering message in attempt to get die-hards to leave

TAKING EVACUATIONS SERIOUSLY

Sources:

[http://
www.nbcnews.co
m/id/9439537/
ns/us_news-
katrina_the_long
_road_back/t/
staying-better-
write-your-ss-
your-body/
#.U6G5oxbePzl,](http://www.nbcnews.com/id/9439537/ns/us_news-katrina_the_long_road_back/t/staying-better-write-your-ss-your-body/#.U6G5oxbePzl)

Wikipedia
Commons

PNAS ARTICLE

Female hurricanes are deadlier than male hurricanes

Kiju Jung^{a,1}, Sharon Shavitt^{a,b,1}, Madhu Viswanathan^{a,c}, and Joseph M. Hilbe^d

Source: <http://www.pnas.org/content/111/24/8782>

Study data: <http://www.pnas.org/content/111/24/8782?tab=ds>

MEDIA RESPONSE

The Opinion Pages | OP-ED COLUMNIST

She Gets No Respect

Sexism Persists, Even Among the Enlightened

JUNE 11, 2014

Scientists discover new source of hurricane danger: Sexism

Do Female-Named Hurricanes Need To Lean In?

by BETH NOVEY

June 03, 2014 1:22 PM ET

People Don't Even Respect Hurricanes with Female Names



Dennis Mersereau

Filed to: SEXISM 6/02/14 4:59pm

7,980 🔥 2 ★

Ladykillers: Hurricanes with female names deadlier



Doyle Rice, USA TODAY

8:06 p.m. EDT June 2, 2014

A NAÏVE START...

What would
'boring' look
like?

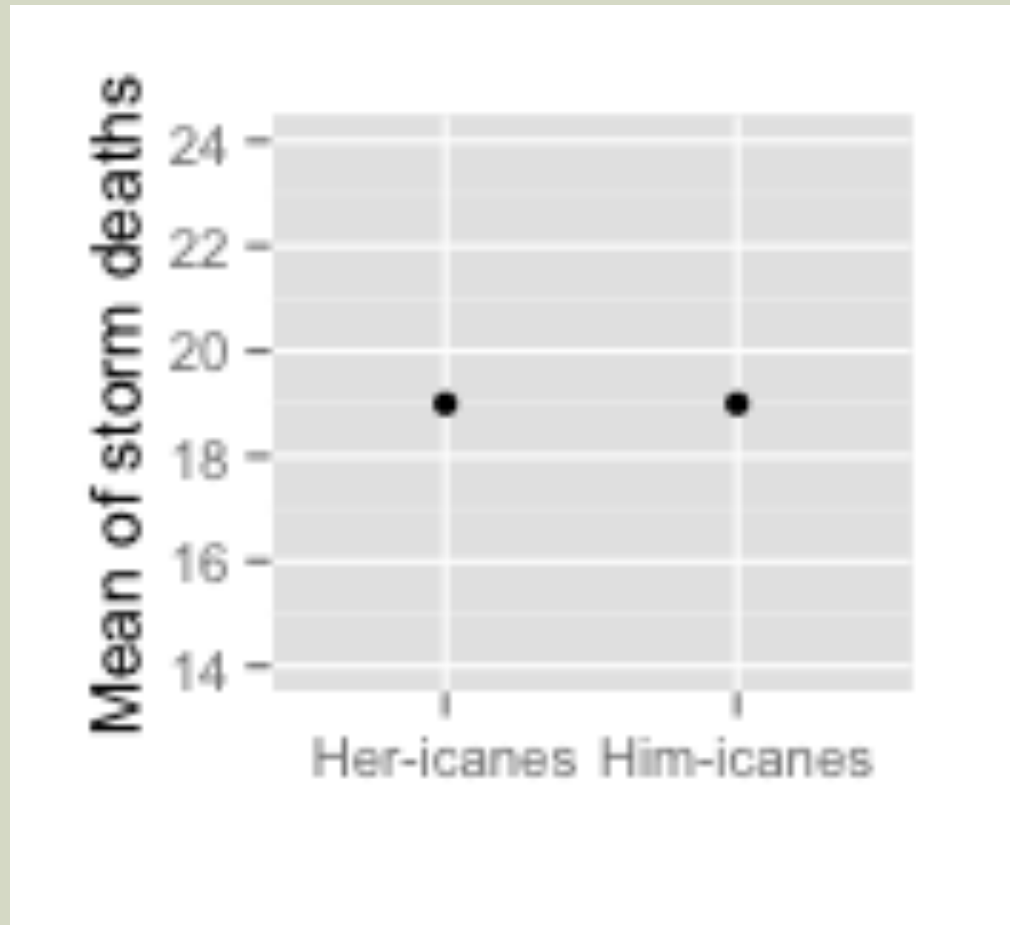


How do the
data differ
from 'boring'?



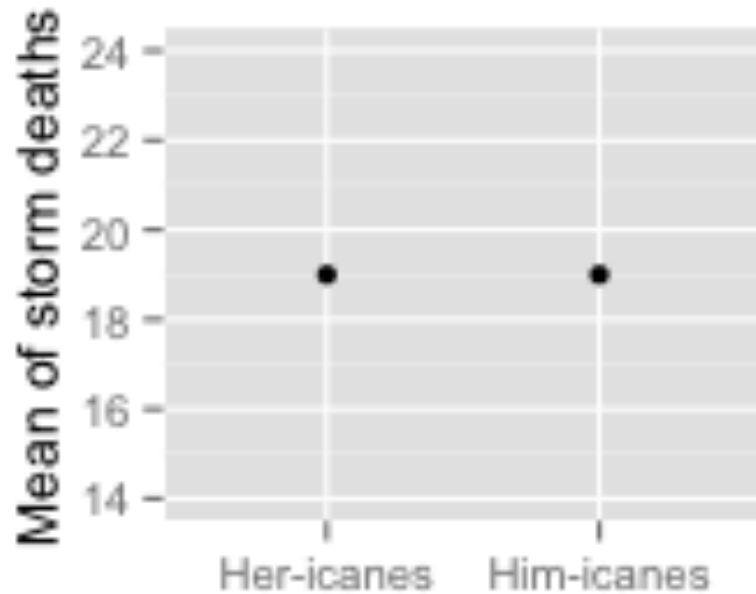
Is this
difference just
by chance?

WHAT WOULD BORING LOOK LIKE?

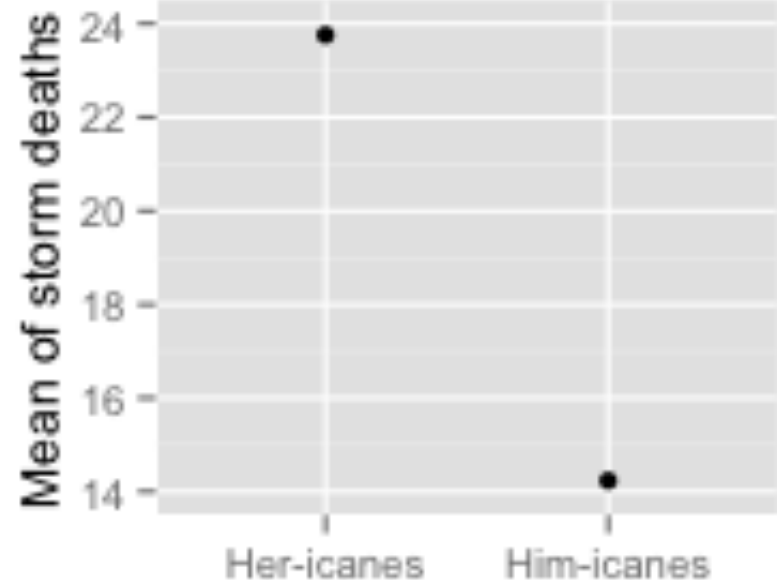


HOW DO THE DATA DIFFER FROM 'BORING'?

Boring



Storm data



IS THIS DIFFERENCE JUST BY CHANCE?



r difference in means

Web

Images

News

Shopping

Videos

More ▾

Search tools

About 375,000,000 results (0.29 seconds)

Population Mean Between Two Independent Samples | R ...

www.r-tutor.com/.../population-mean-between-two-independent-samples ▾

Assuming that the data in mtcars follows the normal distribution, find the 95% confidence interval estimate of the **difference** between the **mean** gas mileage of ...

r - Confidence Interval of Difference of Means between two ...

stackoverflow.com/.../confidence-interval-of-difference-... ▾ Stack Overflow ▾

Jan 20, 2014 - t.test(Bwt ~ Sex, data=cats, conf.level=.99) ...

Kickstarting R - T tests

cran.r-project.org/doc/contrib/Lemon-kickstart/kr_ttest.html ▾ R ▾

brkdn(age~case,infert) 0 1 **Mean** 31.49091 31.53012 Variance 27.60510 ... p- value = 0.956 alternative hypothesis: true **difference in means** is not equal to 0 95 ...

R: Student's t-Test

stat.ethz.ch/R-manual/R-patched/library/stats/.../t.test.html ▾ ETH Zurich ▾

t.test {stats}, R Documentation ... a number indicating the true value of the **mean** (or **difference in means** if you are performing a two sample test). paired. a logical ...

IS THIS DIFFERENCE JUST BY CHANCE?

```
> t.test(alldeaths ~ gender, data = hurr.data, var.equal = TRUE)
```

Two Sample t-test

data: alldeaths by gender

t = 1.0476, df = 90, p-value = 0.2976

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-8.538675 27.588138

sample estimates:

mean in group Her-icanes	mean in group Him-icanes
--------------------------	--------------------------

23.75806	14.23333
----------	----------

Remember: This is the naïve version– we'll talk about the problem with this in a minute.

DID WE MISS SOMETHING?

What would
'boring' look
like?

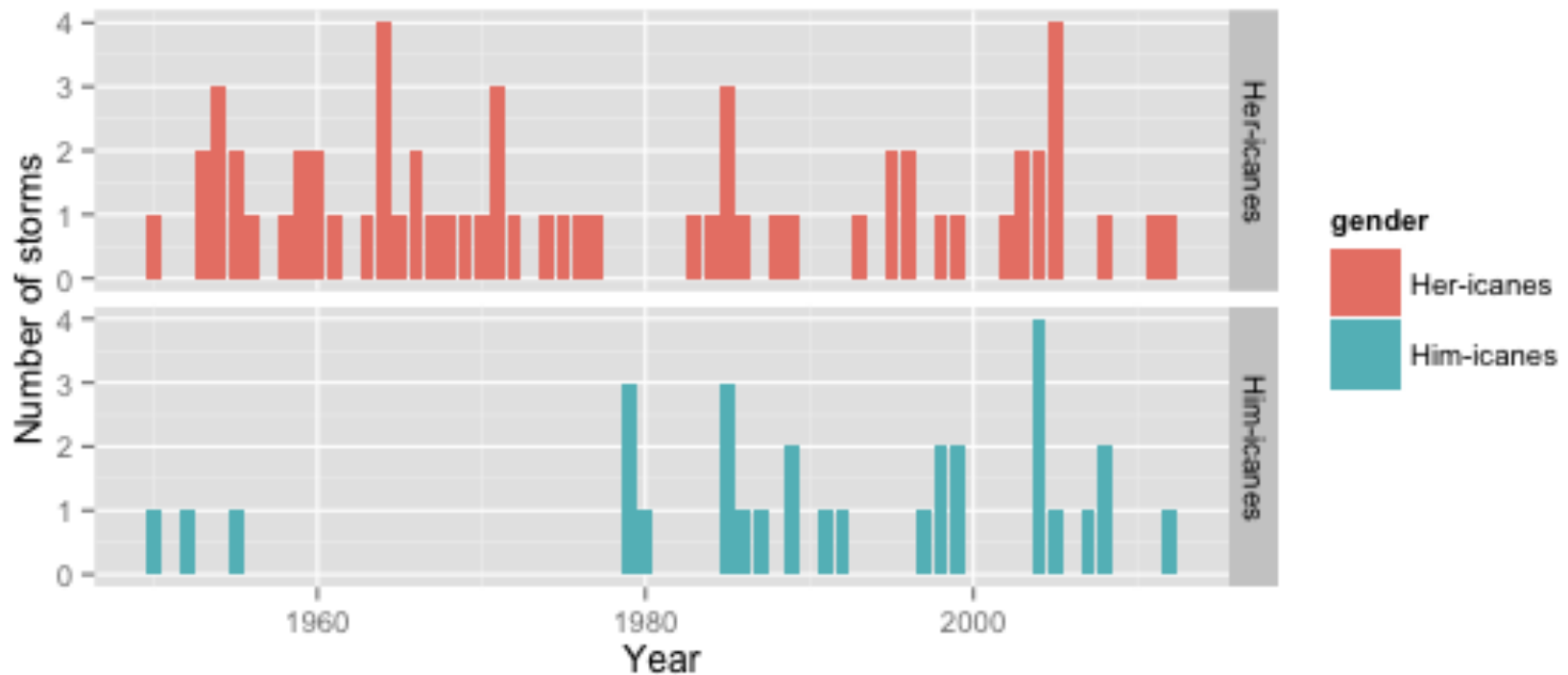


How do the
data differ
from 'boring'?



Is this
difference just
by chance?

STORM GENDER BY TIME



STORM GENDER BY TIME

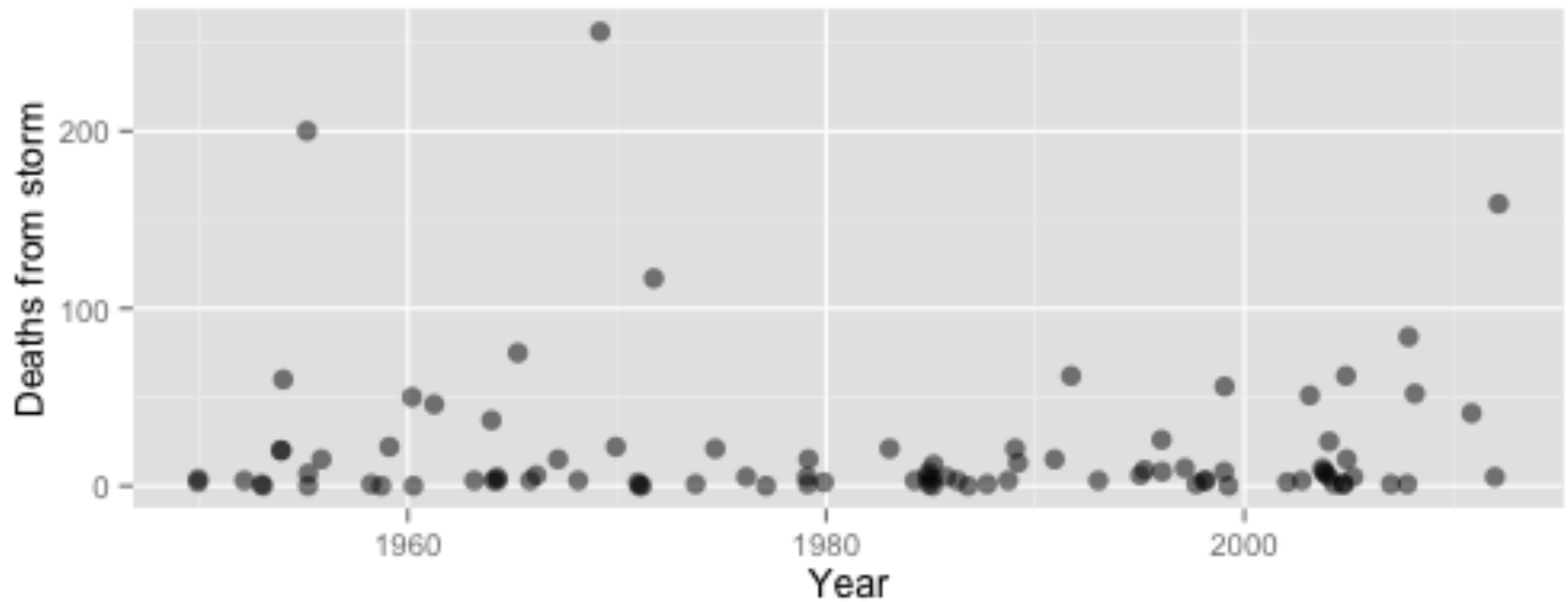
Names used between 1965 – 1979 [\[edit\]](#)

1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
Anna	Alma	Arlene	Abby	Anna	Alma	Arlene	Agnes	Alice	Alma	Amy	Anna	Anita	Amelia	Ana
Betsy	Becky	Beulah	Brenda	Blanche	Becky	Beth	Betty	Brenda	Becky	Blanche	Belle	Babe	Bess	Bob
Carol	Celia	Chloe	Candy	Camille	Celia	Chloe	Carrie	Christine	Carmen	Caroline	Candice	Clara	Cora	Claudette
Debbie	Dorothy	Doria	Dolly	Debbie	Dorothy	Doria	Dawn	Delia	Dolly	Doris	Dottie	Dorothy	Debra	David
Elena	Ella	Edith	Edna	Eve	Ella	Edith		Ellen	Elaine	Eloise	Emmy	Evelyn	Ella	Elena
	Faith	Fern	Frances	Francelia	Felice	Fern		Fran	Fifi	Faye	Frances	Frieda	Flossie	Frederic
	Greta	Ginger	Gladys	Gerda	Greta	Ginger		Gilda	Gertrude	Gladys	Gloria		Greta	Gloria
	Hallie	Heidi		Holly		Heidi				Hallie	Holly		Hope	Henri
	Inez			Inga		Irene							Irma	
	Judith			Jenny		Janice							Juliet	
	Kendra			Kara		Kristy							Kendra	
	Lois			Laurie		Laura								
				Martha										

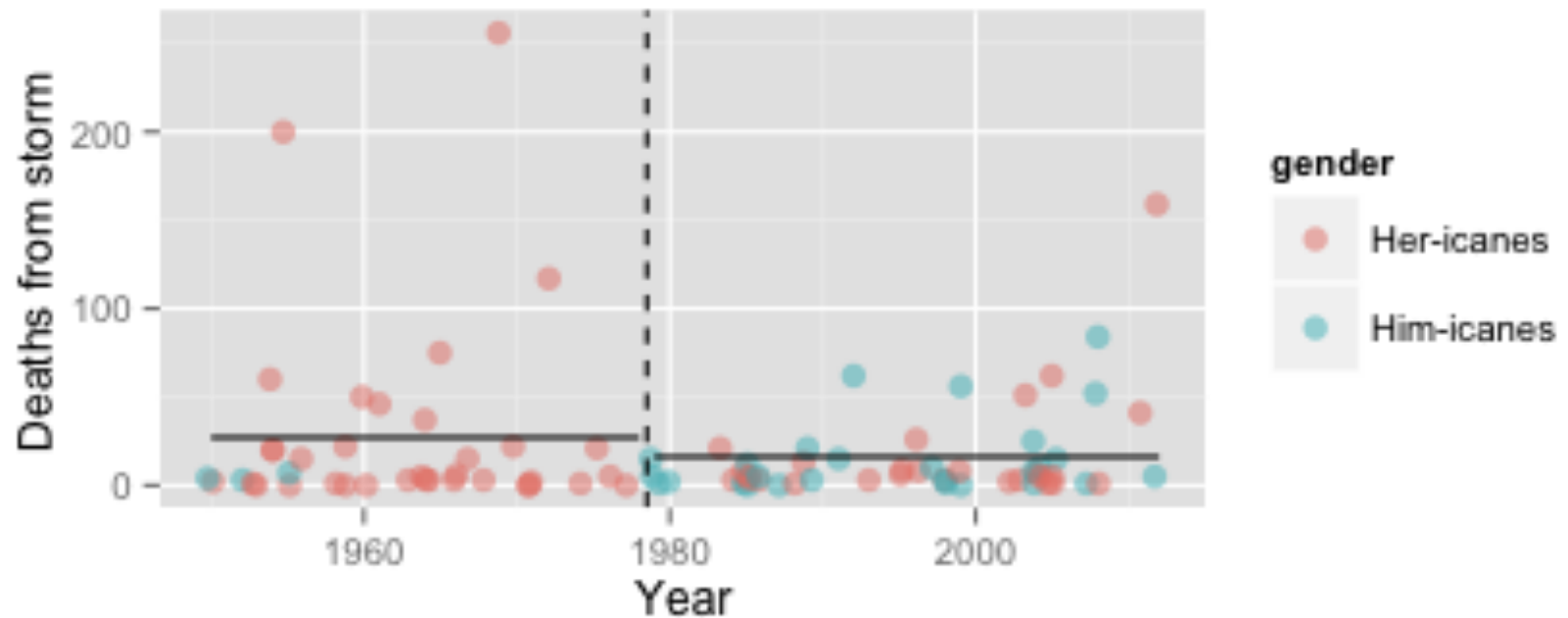
References:^[3]^[8]

Source: Wikipedia

DOES THIS MATTER FOR OUR 'BORING'?

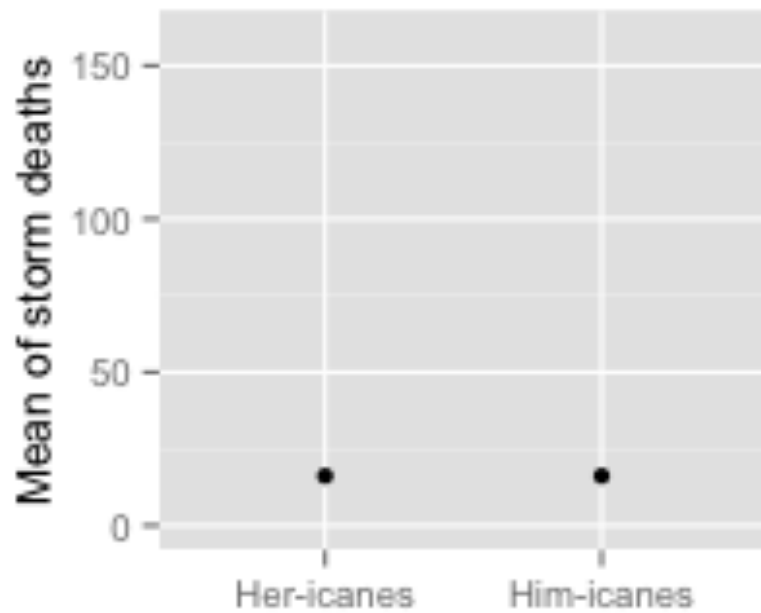


DOES THIS MATTER FOR OUR 'BORING'?

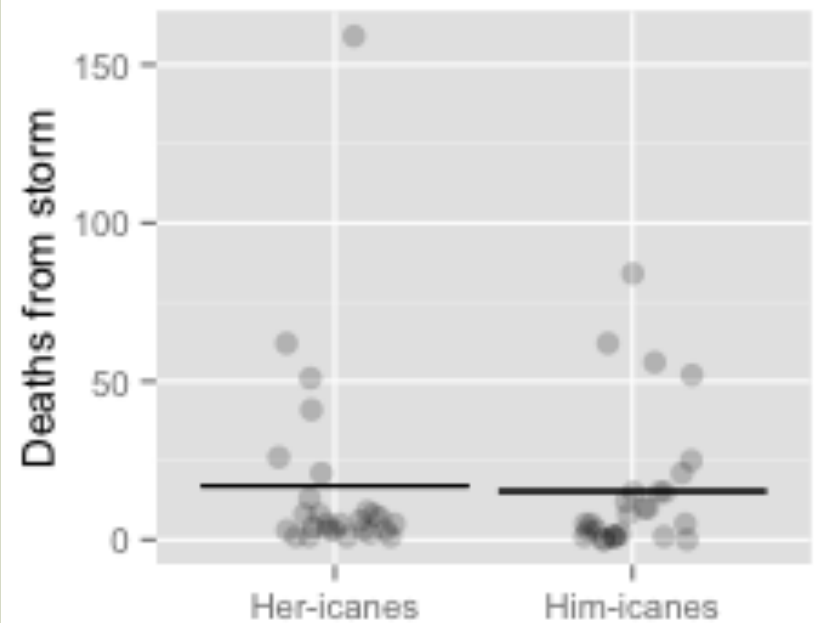


A LESS NAÏVE 'BORING'

Boring



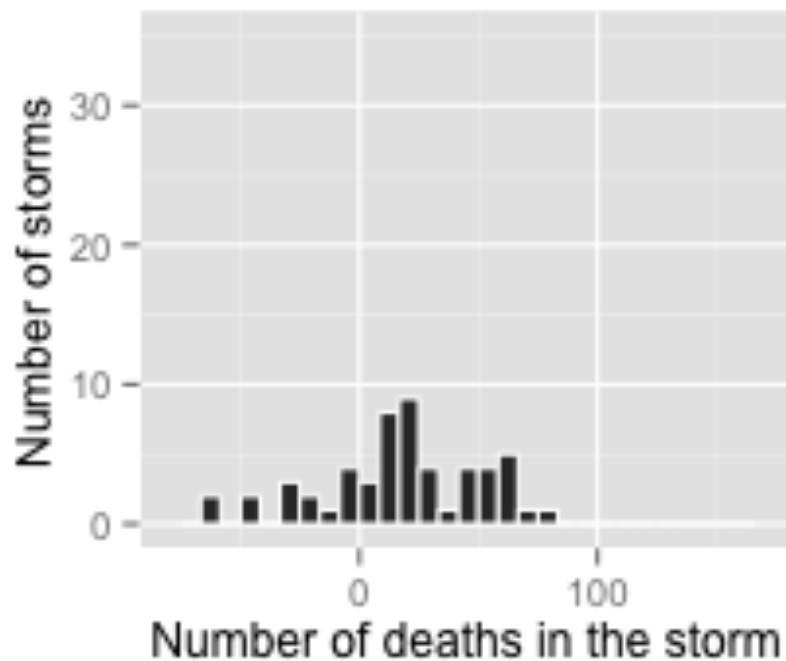
Our data



***Limited* to storms in 1979 or later!!**

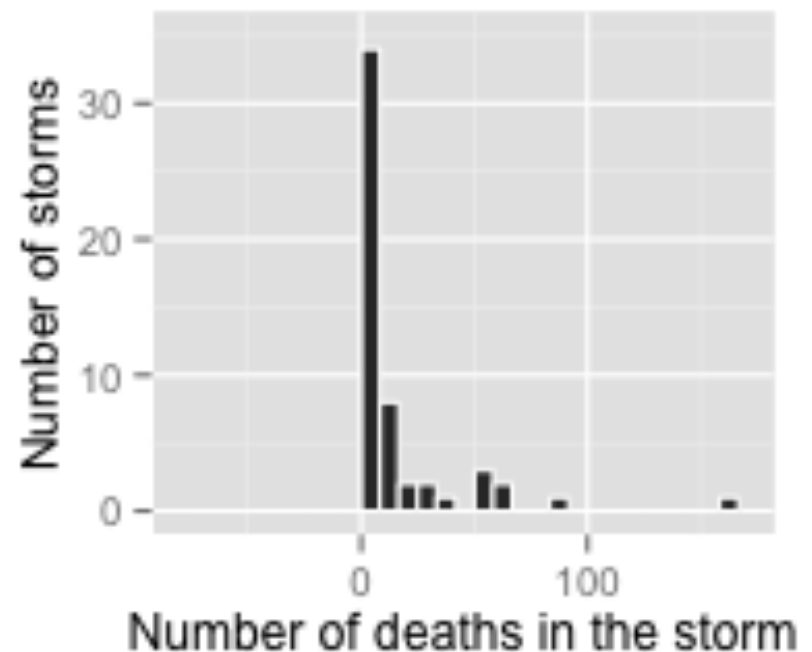
REFINING 'BORING' EVEN MORE (TO MAKE IT EASIER TO ANALYZE CHANCE)

Simulated distribution
(normal)



(mean = 16; sd = 27.5)

Distribution of hurricane
deaths



T-TEST OF TWO MEANS

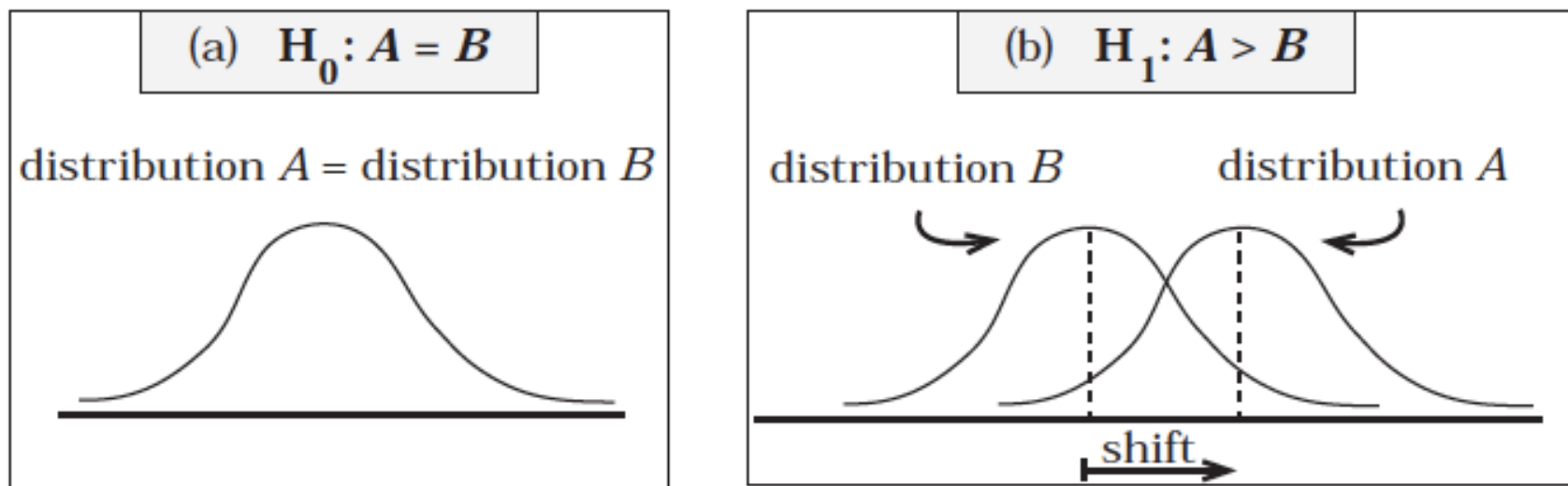


Figure 2 : Illustration of $H_0 : A = B$ versus $H_1 : A > B$.

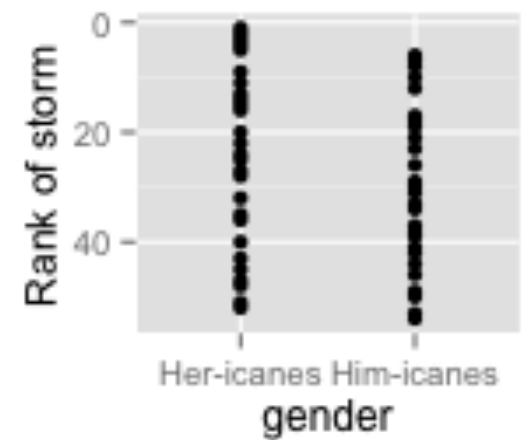
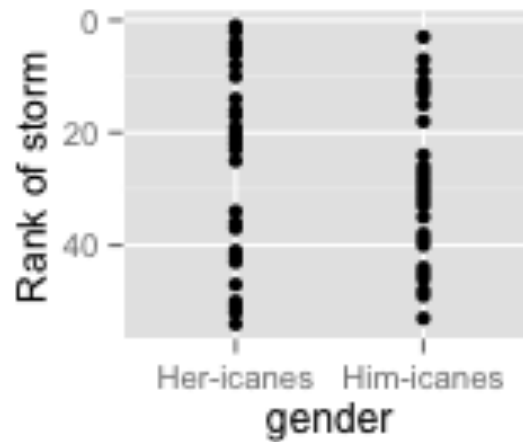
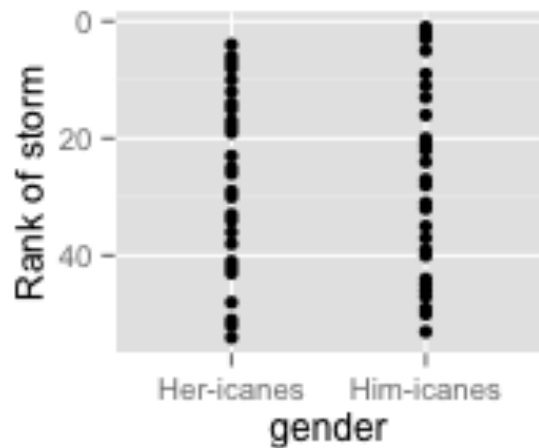
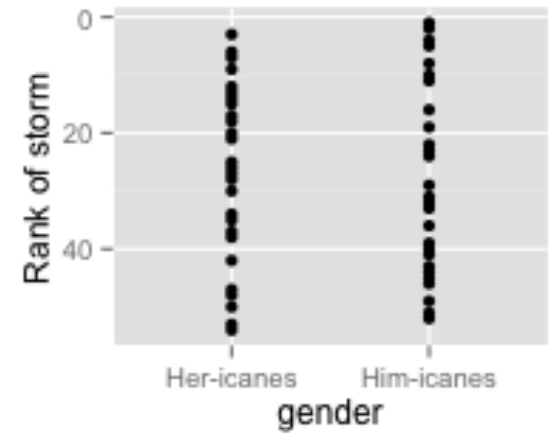
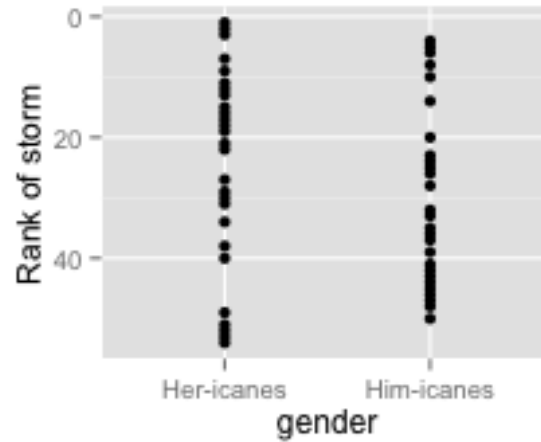
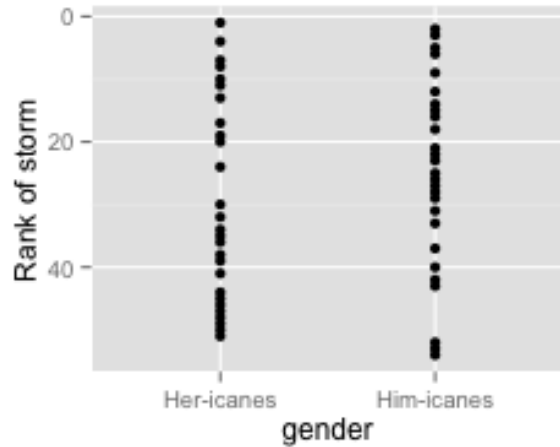
Source: <https://www.stat.auckland.ac.nz/~wild/ChanceEnc/Ch10.wilcoxon.pdf>

NEW BORING: THINKING BY RANKS, NOT MEANS

```
> head(new.df[,c("Year", "Name", "alldeaths", "rank")])
```

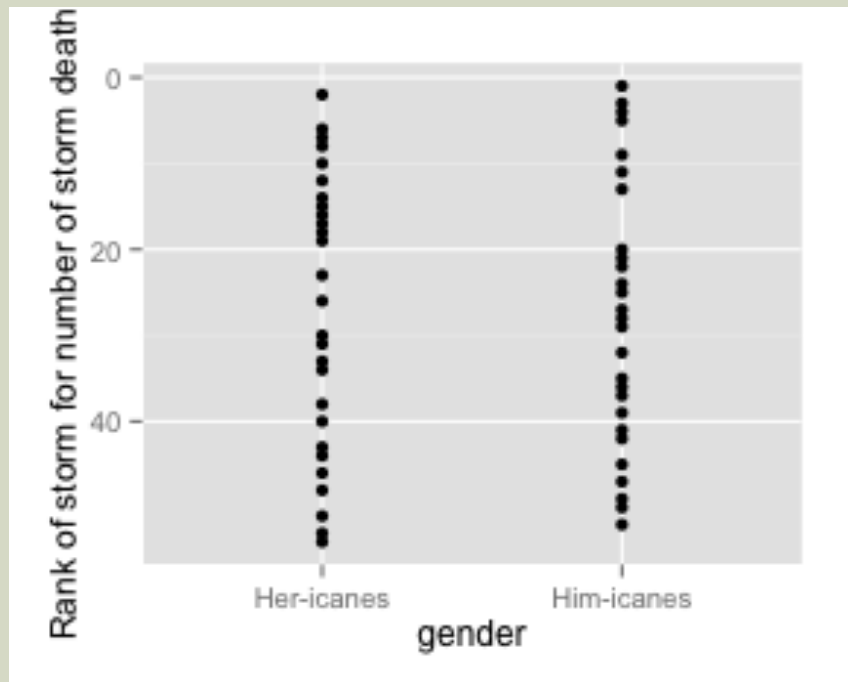
	Year	Name	alldeaths	rank
92	2012	Sandy	159	1
89	2008	Ike	84	2
59	1992	Andrew	62	3
84	2005	Rita	62	4
70	1999	Floyd	56	5
88	2008	Gustav	52	6

A BETTER 'BORING'

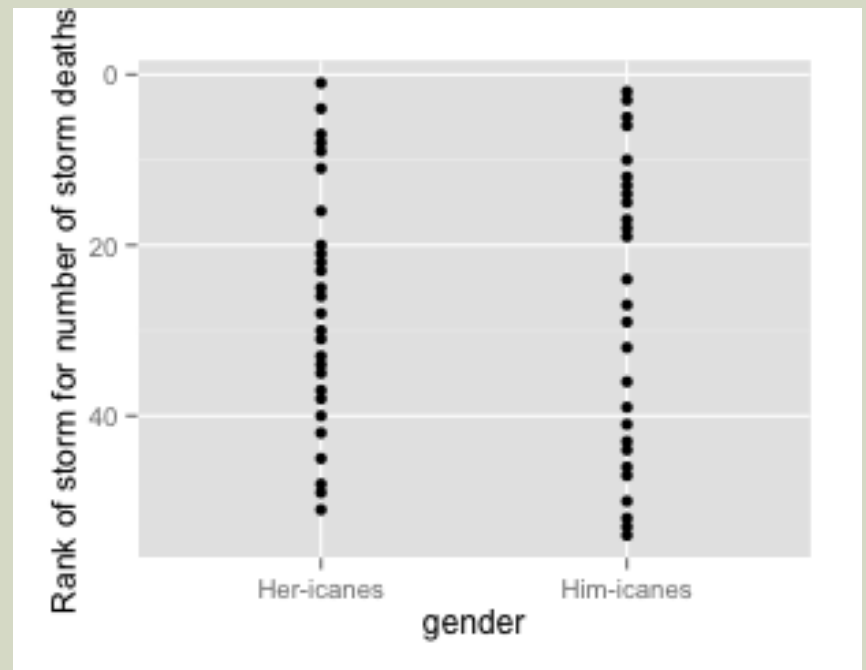


A BETTER 'BORING'

Boring (One simulation)



Our data



TESTING IF DIFFERENCE BETWEEN DATA AND 'BORING' IS CHANCE

```
> wilcox_test(alldeaths ~ gender, data = new.df, conf.int = TRUE)
```

Asymptotic Wilcoxon Mann-Whitney Rank Sum Test

data: alldeaths by gender (Her-icanes, Him-icanes)

Z = 0.2085, p-value = 0.8348

alternative hypothesis: true mu is not equal to 0

95 percent confidence interval:

-4.999893 3.000041

sample estimates:

difference in location

6.576595e-05

BLOG POSTS THAT GO OVER THE ANALYSIS

- http://www.theguardian.com/science/grrlscientist/2014/jun/04/hurricane-gender-name-bias-sexism-statistics?CMP=tw_tfd
- <http://phenomena.nationalgeographic.com/2014/06/02/why-have-female-hurricanes-killed-more-people-than-male-ones/>
- <http://mashable.com/2014/06/02/hurricanes-female-names/>
- http://www.slate.com/blogs/future_tense/2014/06/03/are_hurricanes_named_after_women_more_dangerous_not_so_fast.html
- <http://blog.minitab.com/blog/the-statistics-game/do-the-data-really-say-female-named-hurricanes-are-more-deadly>

DATA SCIENCE COMPETITIONS

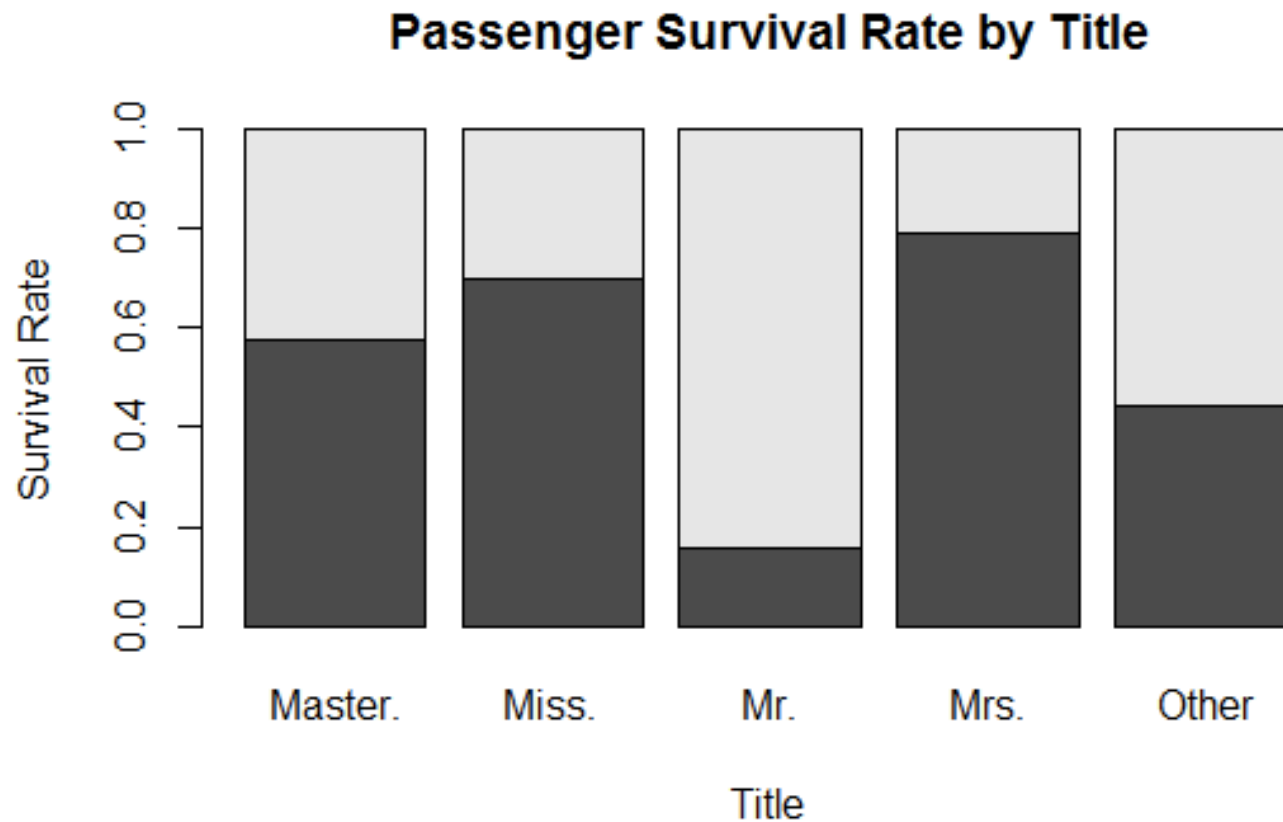
ADHD 200 PREDICTION COMPETITION

Johns Hopkins Team Wins 'ADHD-200 Global Competition' to Develop Imaging-Based Psychiatric Diagnosis Protocol

Find out more:

<http://www.smart-stats.org/content/johns-hopkins-team-wins-'adhd-200-global-competition'-develop-imaging-based-psychiatric>

KAGGLE COMPETITIONS



<https://www.kaggle.com/competitions>

NETFLIX PRIZE

The logo features the word "NETFLIX" in a bold, white, sans-serif font with a black outline. It is positioned on a red rectangular background. Below this, a yellow rectangular area contains the words "Netflix Prize" in a white, rounded, sans-serif font with a black outline.

Netflix Prize

Read more: http://en.wikipedia.org/wiki/Netflix_Prize

CURRENT EVENTS: DATA SCIENCE EDITION

- Slate.com article: Rappers vs. Scotus: Who uses a bigger vocabulary, Jay-Z or Scalia
http://www.slate.com/articles/news_and_politics/view_from_chicago/2014/06/supreme_court_and_rappers_who_uses_a_bigger_vocabulary_jay_z_or_scalia.html
- When will the next 'Game of Thrones' book come out?
<http://fivethirtyeight.com/features/predicting-when-game-of-thrones-author-george-r-r-martin-will-come-out-with-his-next-book/>
- Analyzing the Spurs' win:
<http://fivethirtyeight.com/datalab/the-spurs-were-an-outlier-of-unselfishness/>