

# COMP10020

## Introduction to Programming II

### **Data Science Case Studies**

Dr. Brian Mac Namee

[brian.macnamee@ucd.ie](mailto:brian.macnamee@ucd.ie)

School of Computer Science

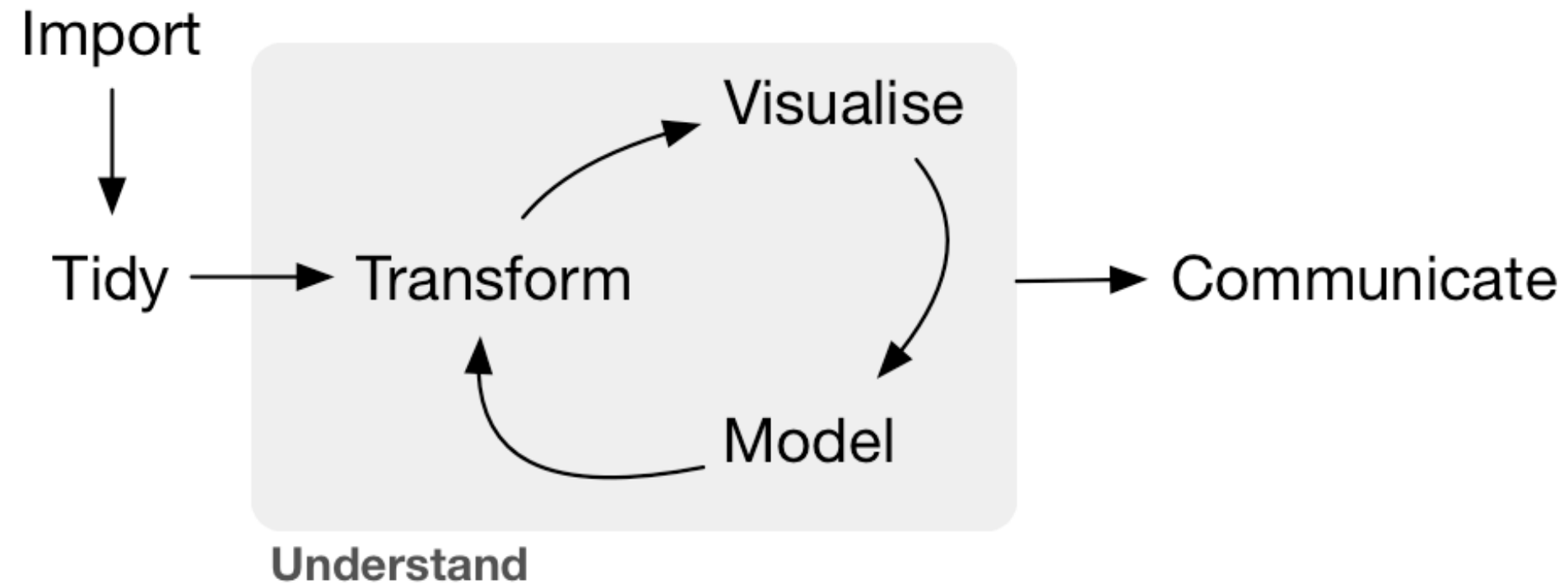
University College Dublin

# Data Science Case Studies

---

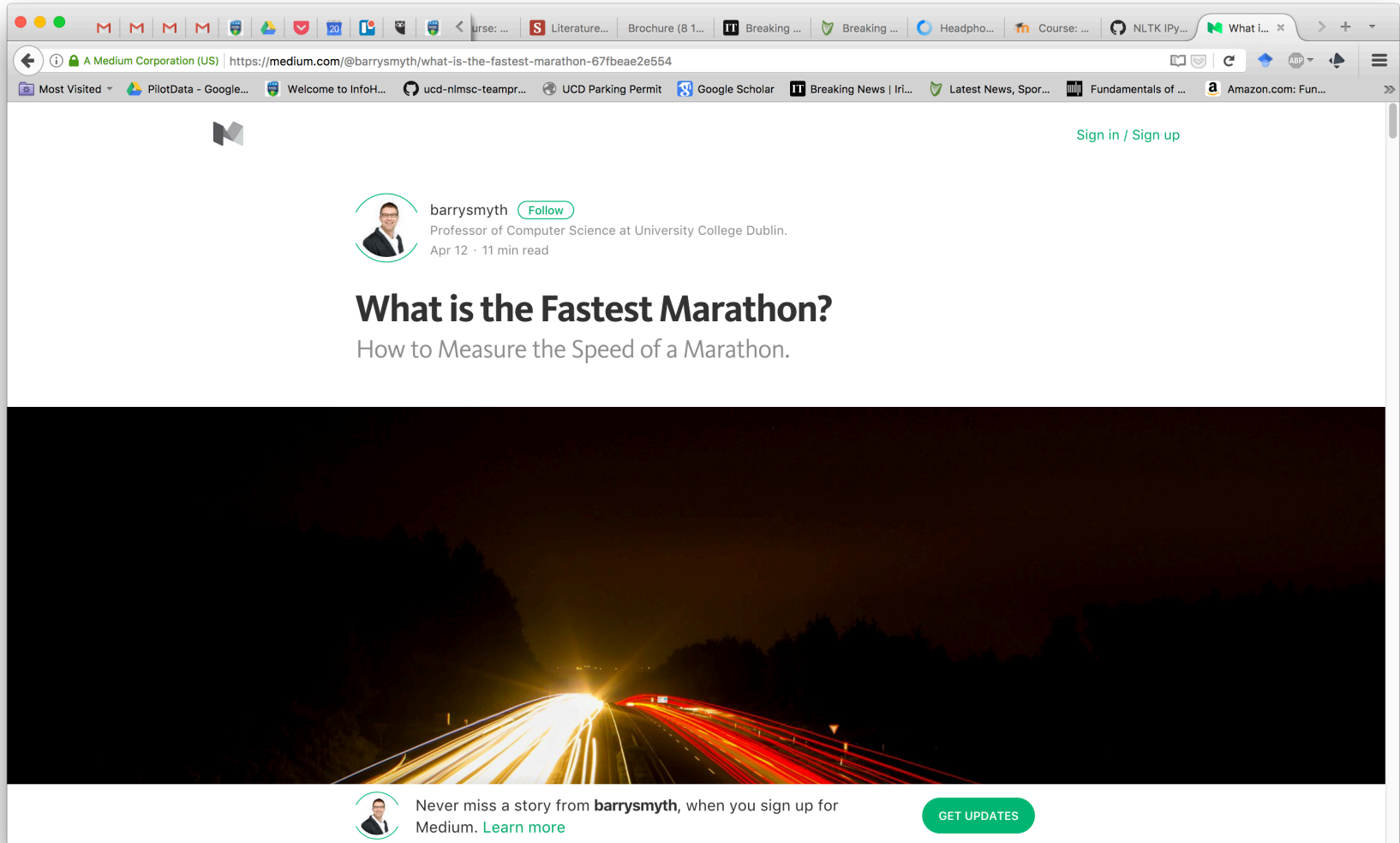
In this lecture we will look at 3 key data science case studies illustrating everything that we have been talking about

- What Is The Fastest Marathon?
- Which Songs Are Timeless?
- Analysing Donal Trump's Tweets

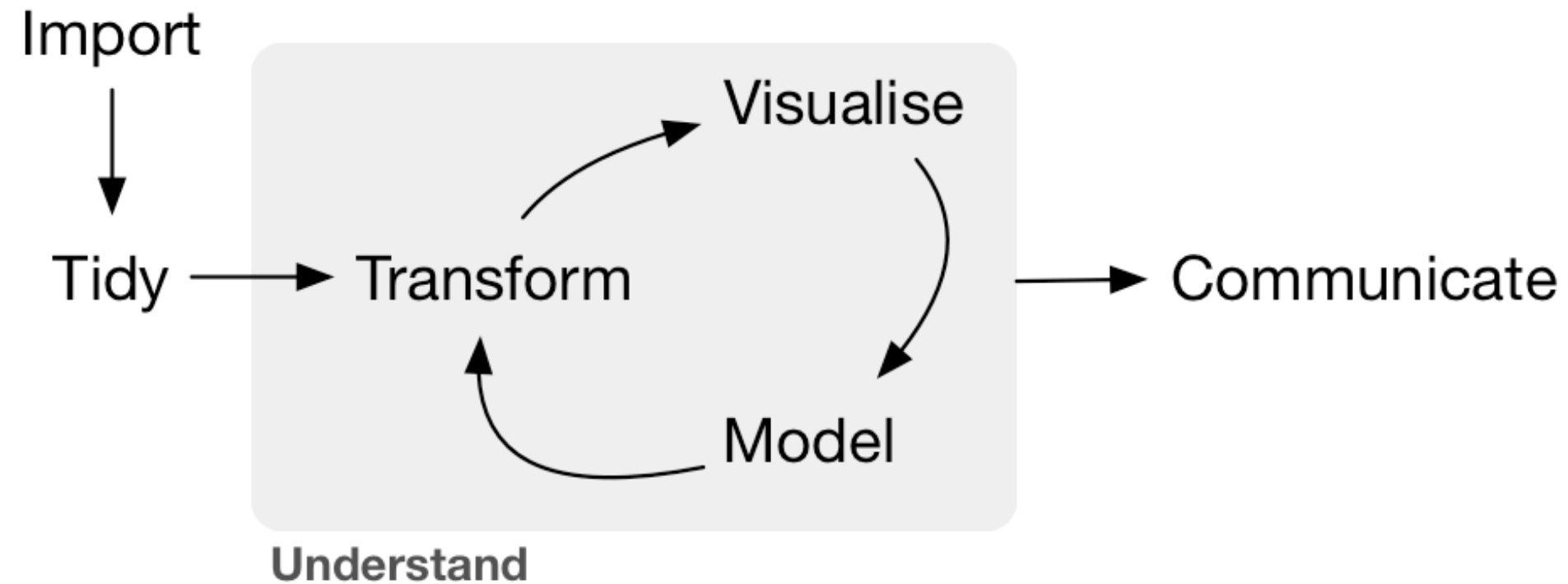


# **WHAT IS THE FASTEST MARATHON?**

# What Is The Fastest Marathon?

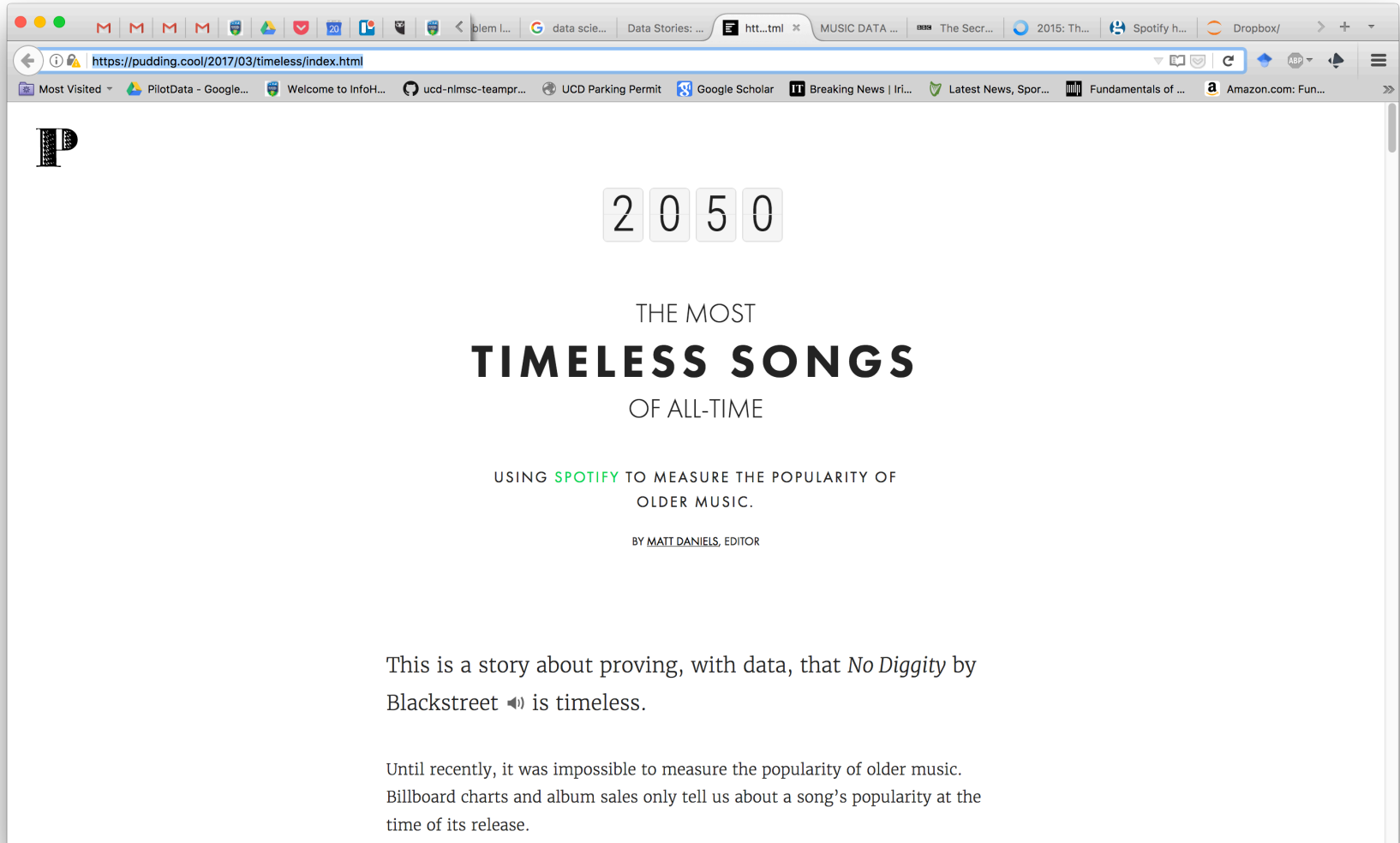


<https://medium.com/@barrysmth/what-is-the-fastest-marathon-67fbae2e554>



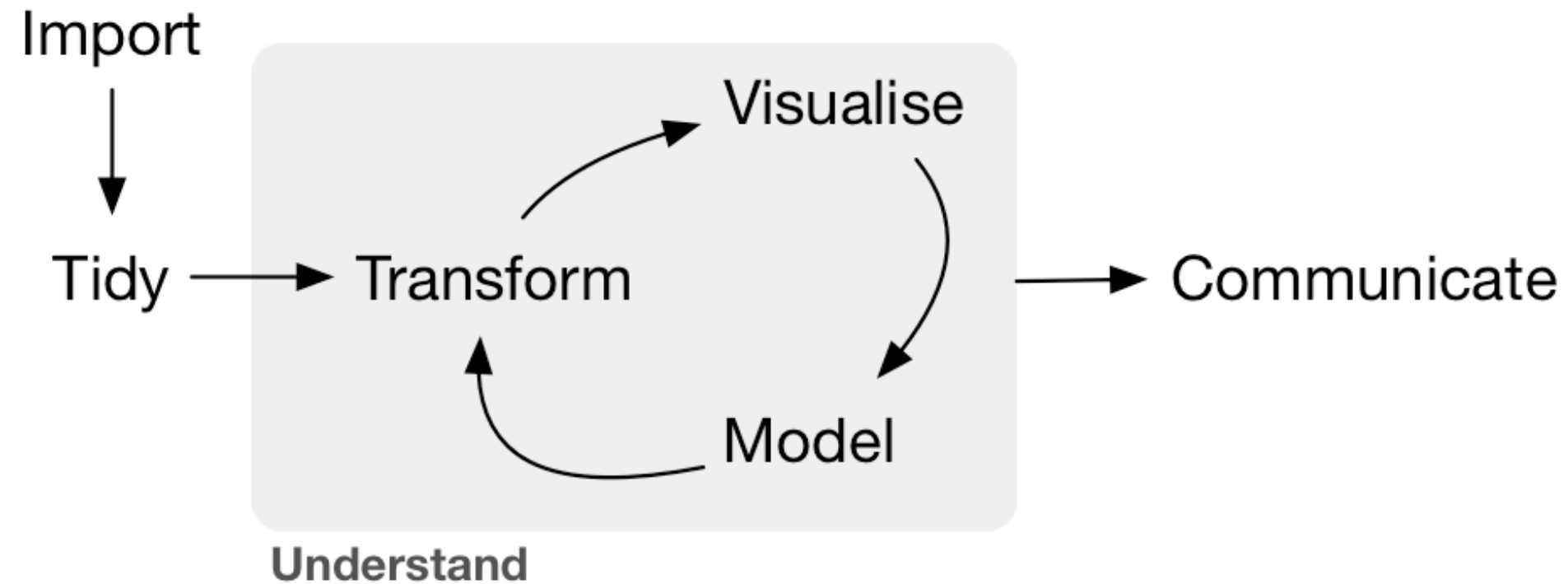
**WHICH SONGS ARE  
TIMELESS?**

# Which Songs Are Timeless?



<https://pudding.cool/2017/03/timeless/index.html>





# **ANALYSING DONAL TRUMP'S TWEETS**

# Analysing Donal Trump's Tweets

The screenshot shows a web browser window with the URL [varianceexplained.org/r/trump-tweets/](http://varianceexplained.org/r/trump-tweets/). The page features a navigation bar with links: VARIANCE EXPLAINED, ABOUT ME, POSTS, R COURSE, and INTRODUCTION TO EMPIRICAL BAYES. On the left, there is a profile for David Robinson, a Data Scientist at Stack Overflow, with links to his email, Twitter, GitHub, and Stack Overflow. Below his profile is a 'Subscribe' section with a text input for 'Your email' and a 'Subscribe to this blog' button. Further down are 'Recommended Blogs' including DataCamp, R Bloggers, RStudio Blog, and R4Stats. The main content area displays the article title 'Text analysis of Trump's tweets confirms he writes only the (angrier) Android half' and a paragraph of text. Below the text are two side-by-side images of tweets from Donald J. Trump. The left tweet, dated Aug 5, 2016, at 8:59 PM, says 'Good luck #OpeningCeremony pic.twitter.com/...' and has 27,391 Likes. The right tweet, dated Aug 6, 2016, at 11:11 AM, says 'Heading to talking about SHORT CIRCUIT' and has 4,451 Likes. Below these is a tweet from Todd Vaziri (@tvaziri) dated 8:20 PM - 6 Aug 2016, which contains the text: 'Every non-hyperbolic tweet is from iPhone (his staff). Every hyperbolic tweet is from Android (from him)'. This tweet has 10,468 retweets and 14,846 likes.

VARIANCE EXPLAINED ABOUT ME POSTS R COURSE INTRODUCTION TO EMPIRICAL BAYES

**Text analysis of Trump's tweets confirms he writes only the (angrier) Android half**

I don't normally post about politics (I'm not particularly savvy about polling, which is where data science has had the largest impact on politics). But this weekend I saw a hypothesis about Donald Trump's twitter account that simply begged to be investigated with data:

**Donald J. Tru**  
Good luck #  
#OpeningCe  
pic.twitter.c

27,391 Likes  
Aug 5, 2016 at 8:59 PM

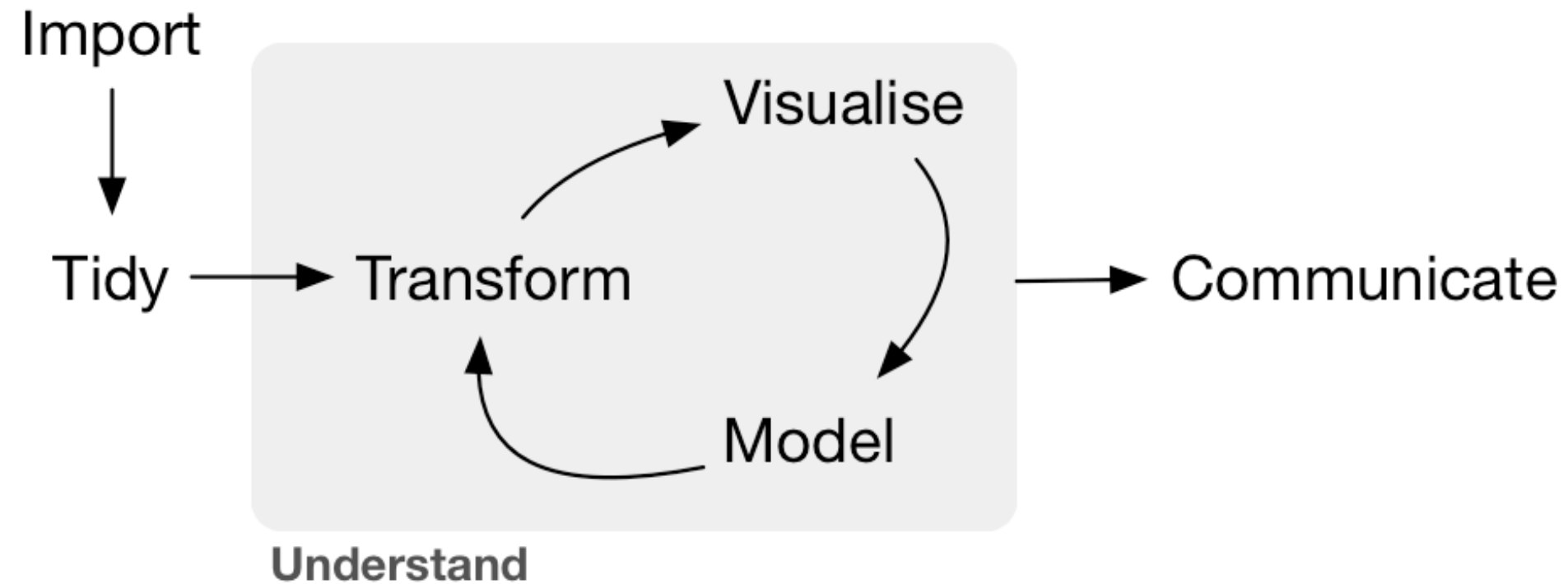
**Donald J. Tr**  
Heading to  
talking abo  
SHORT CIR

4,451 Likes  
Aug 6, 2016 at 11:11 AM

**Todd Vaziri** @tvaziri  
Every non-hyperbolic tweet is from iPhone (his staff).  
Every hyperbolic tweet is from Android (from him).  
8:20 PM - 6 Aug 2016  
10,468 retweets 14,846 likes

**David Robinson**  
Data Scientist at Stack Overflow, works in R and Python.  
Email Twitter GitHub Stack Overflow  
Subscribe  
Your email  
Subscribe to this blog  
Recommended Blogs  
DataCamp  
R Bloggers  
RStudio Blog  
R4Stats

<http://varianceexplained.org/r/trump-tweets/>

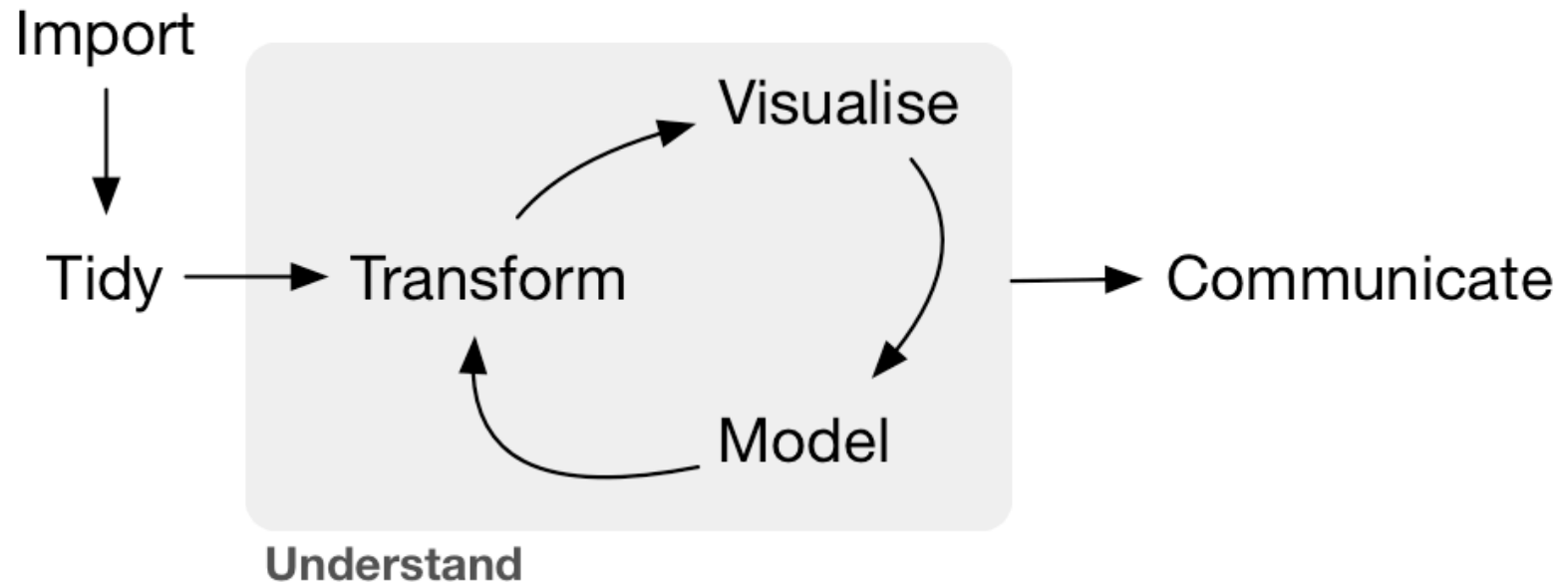


# SUMMARY

# Summary

---

In this lecture we have seen how the same basic approach to data science projects can be used for very different challenges



<https://ismayc.github.io/moderndiver-book/2-intro.html>