

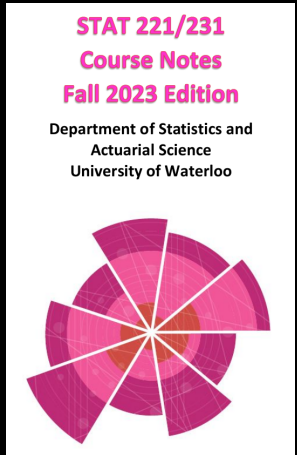
Keeping it Real: Data Analysis Assignments in STAT 231 Using the Stanford Open Policing Project

Michael Wallace
michael.wallace@uwaterloo.ca

Assignment dataset downloader (more on this later!):
`shiny.math.uwaterloo.ca/sas/stat231/stopdownloader/`

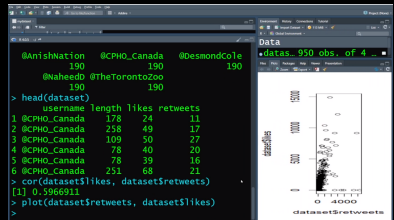
Slide deck and links available at:
`https://mpwallace.github.io/`

- Intro stats course (follows STAT 230)
- Enrolment: 400-700 (multi-section)
- Wide range of abilities/backgrounds
- Challenge: student engagement, especially “why is this useful?”



Introduce/expand R experience through assignments:

- Prior experience: little/none
- Relatively basic commands/analyses
- Not a coding/programming course!
- Focus: interpretation of results



R material primarily covered through tutorial videos

- Realistic data
- Realistic analysis questions
- Broad range of analyses
- Unique datasets per student



Stanford Open Policing Project

**THE STANFORD
OPEN POLICING
PROJECT**

On a typical day in the United States, police officers make more than 50,000 traffic stops. Our team is gathering, analyzing, and releasing records from millions of traffic stops by law enforcement agencies across the country. Our goal is to help researchers, journalists, and policymakers investigate and improve interactions between police and the public.

[VIEW DATA](#)

<https://openpolicing.stanford.edu/>

Stanford Open Policing Project

State	Download	Stops	Time Range	Stop Date	Stop Time	Stop Location	Driver Race	Driver Sex	Driver Age	Search Conducted	Contraband Found	Citation Issued	Warning Issued	Frisk Performed	Arrest Made	Reason for Stop	Violation
AR																	
Little Rock	R	13,641	Dec 2016 - Nov 2017	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>							
AZ																	
Gilbert	R	480,599 ²	Dec 2007 - May 2018	<input type="checkbox"/>	<input type="checkbox"/>												
Mesa	R	157,015 ²	Dec 2013 - Sep 2019	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
State Patrol	R	3,494,153 ²	Jan 2009 - Dec 2017	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
CA																	
Anaheim	R	87,876	Dec 2011 - Mar 2017	<input type="checkbox"/>											<input type="checkbox"/>		
Bakersfield	R	189,685	Mar 2008 - Mar 2018	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>						
Long Beach	R	365,924 ²	Dec 2007 - Dec 2017	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>						<input type="checkbox"/>
Los Angeles	R	5,418,400 ²	Dec 2009 - Jun 2018	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>									
Oakland	R	133,405 ²	Mar 2013 - Dec 2017	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<https://openpolicing.stanford.edu/data/>

Note: Available data varies by location.

I chose Chicago and San Francisco and the following variates:

Stop data	Time, Date, Location (lat/long)
Subject demographics	Age, Race, Sex, Vehicle make
Offence data	Violation, Outcome

Pre-processing steps:

- Time: Converted to hours after midnight
- Date: Converted to day of week
- Vehicle make: Collapsed to Toyota, Ford, Chevrolet, Honda, Other

Main data issues:

- Missing data:
 - Complete case analysis (\approx 2 million traffic stops)
- Sex:
 - Coded as Female or Male only
- Race:
 - San Francisco: Asian/Pacific Islander, Black, Hispanic, White, Other
 - Chicago: Black, Hispanic, White

← → ↻ 🔒 shiny.math.uwaterloo.ca/sas/stat231/stopdownloader/

STAT 231 Traffic Stop Dataset Downloader

Enter the password (found on LEARN) and your student ID number. A 'Download' button will appear. Click the button to download your sample from the primary dataset. Once your sample is downloaded, please upload it to the LEARN dropbox *immediately*! :)

Enter password:

Enter your ID number:

Click ONCE and wait, it may take a few moments!

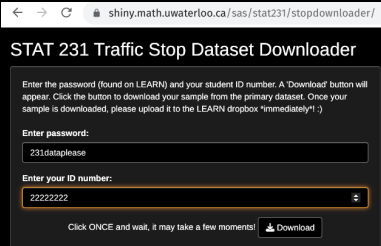
 Download

<https://shiny.math.uwaterloo.ca/sas/stat231/stopdownloader/>

Random sample of ≈ 800 entries, tied to student ID number.

Generating Unique Datasets

- Datasets are unique
- But not *too* unique!
- Tied to student ID
- Uploaded to course website for pre-checks



The screenshot shows a web browser window with the address bar containing the URL `shiny.math.uwaterloo.ca/sas/stat231/stopdownloader/`. The page title is "STAT 231 Traffic Stop Dataset Downloader". The main content area has a dark background and contains the following text and form elements:

Enter the password (found on LEARN) and your student ID number. A 'Download' button will appear. Click the button to download your sample from the primary dataset. Once your sample is downloaded, please upload it to the LEARN dropbox "immediately"! :)

Enter password:

Enter your ID number:

Click ONCE and wait, it may take a few moments!

Assignment structure:

- 2-4 'Analyses', each focused on 1-2 variates
- Sub-parts of R analysis followed by interpretation, e.g.:
 - (a) Generate a bar plot of Subject Race or Vehicle Make stratified by Subject Sex
 - (b) Discuss any similarities/differences
- Students are given some choice (e.g., to analyze Race or Vehicle Make)
- Students submit a structured 'Report' and their R code

- TAs verify numerical/graphical results using R code
- TAs review and give feedback on interpretation questions
- 'Debrief' document containing example analysis and highlighting key learning objectives also provided



The good:

- Good engagement/positive student feedback
- Opportunity to discuss broader issues/topics
- Relatively easy grading

The not-so-good:

- Non-trivial setup and pre-processing
- Some analyses a little contrived
- Extra care needed to avoid teaching 'bad habits'

- **Stanford Open Policing Project**
<https://openpolicing.stanford.edu/>
- **Sample Downloader** <https://shiny.math.uwaterloo.ca/sas/stat231/stopdownloader/>
Password: 231dataplease; ID number $\in [2 \times 10^7, 3 \times 10^7)$
- **My website:** <https://mpwallace.github.io/>
- **My email:** michael.wallace@uwaterloo.ca
- **My socials:** @statacake