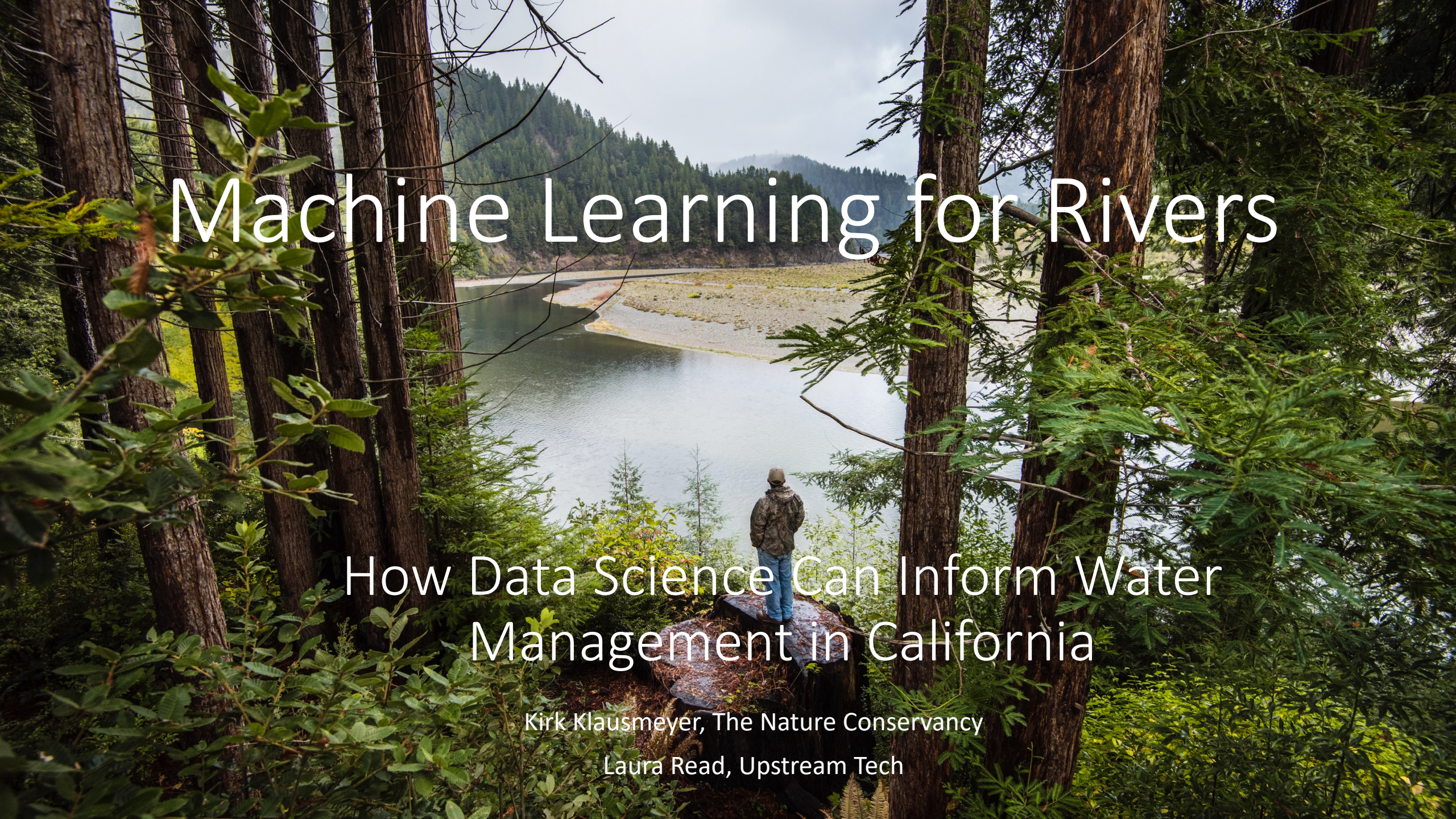


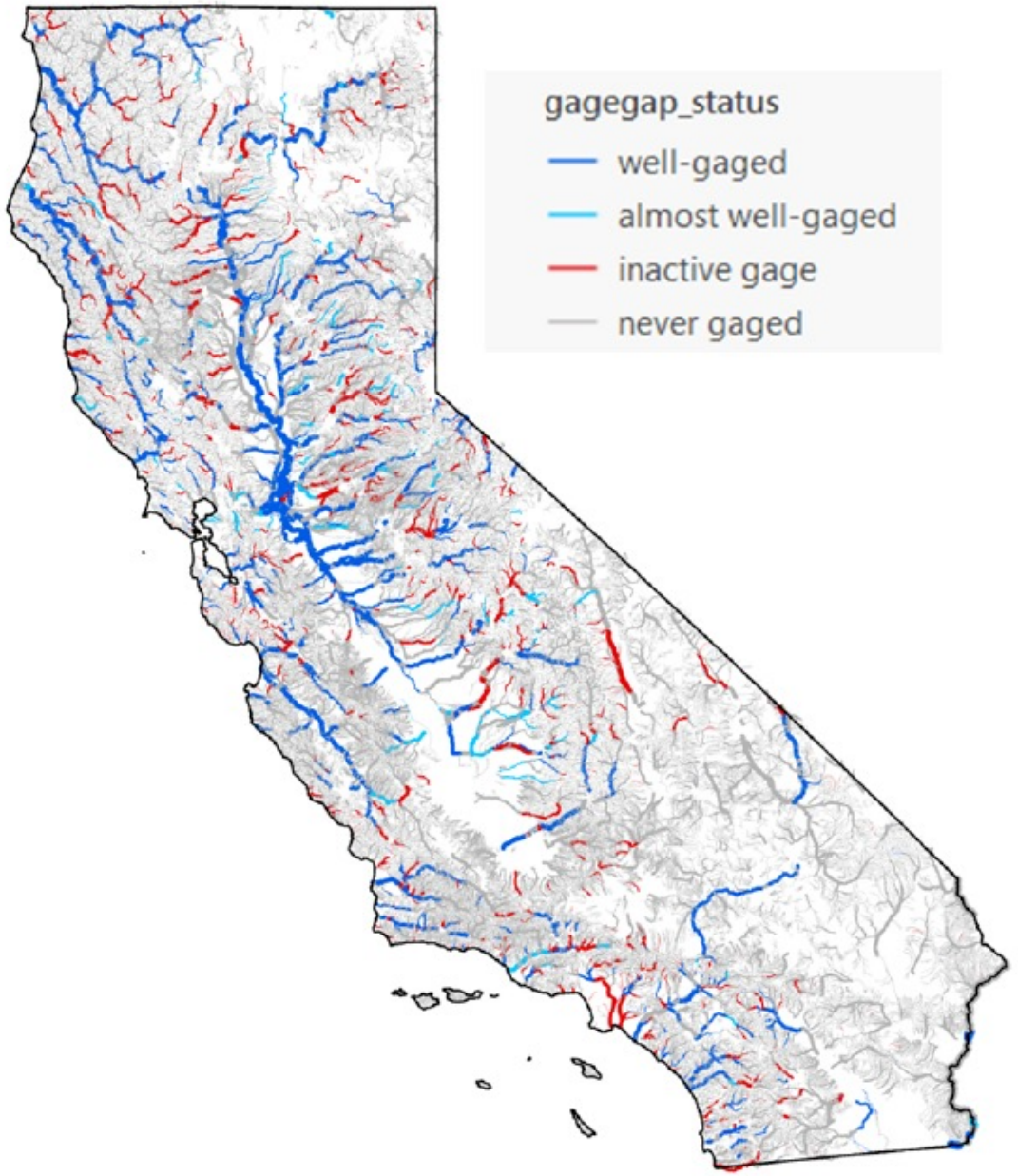
Machine Learning for Rivers

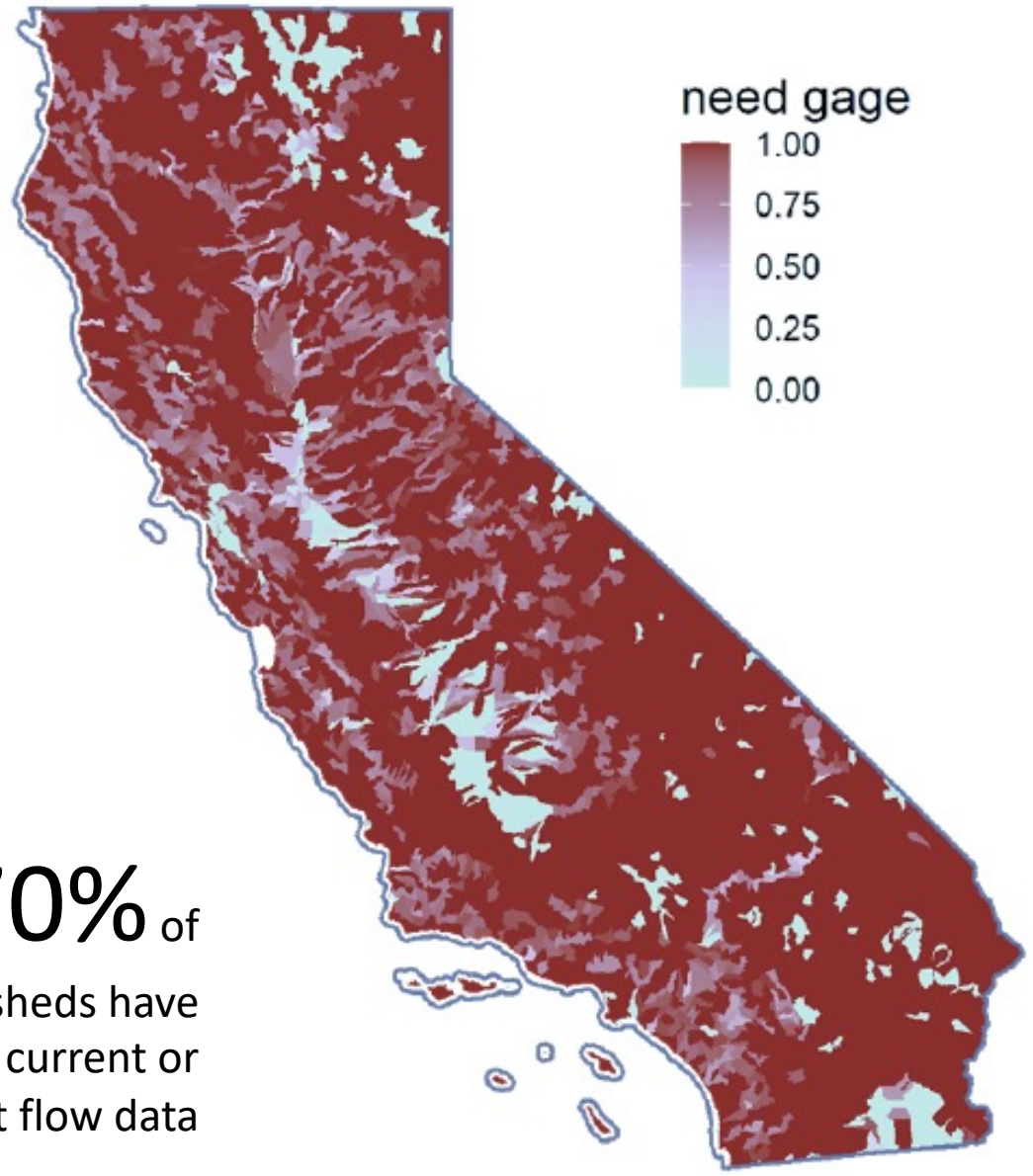
How Data Science Can Inform Water Management in California

Kirk Klausmeyer, The Nature Conservancy

Laura Read, Upstream Tech

The Problem





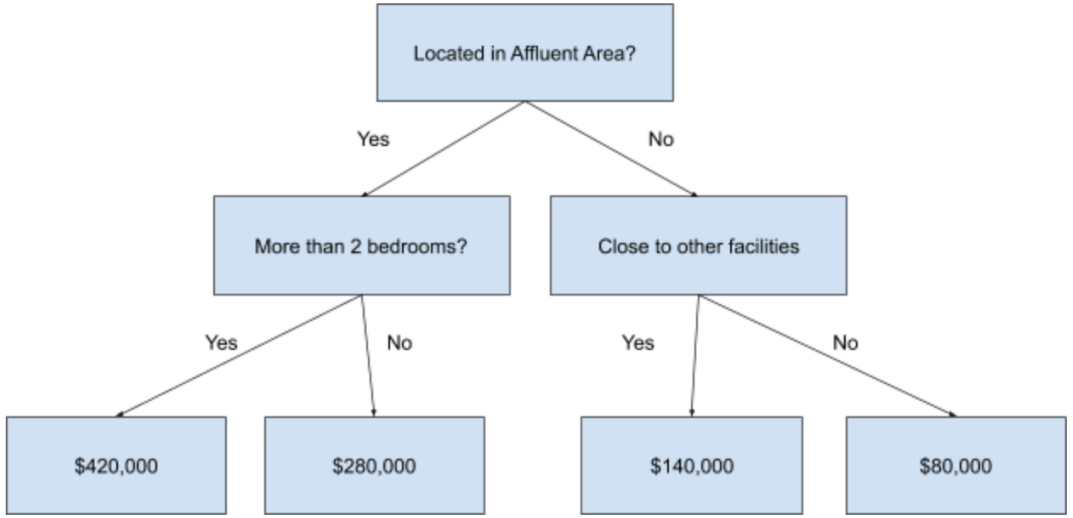
70% of watersheds have no current or past flow data

Source: CALIFORNIA STREAM GAGING PRIORITIZATION PLAN 2022. Prepared by: Department of Water Resources, State Water Resources Control Board, Department of Fish and Wildlife, Department of Conservation - California Geological Survey

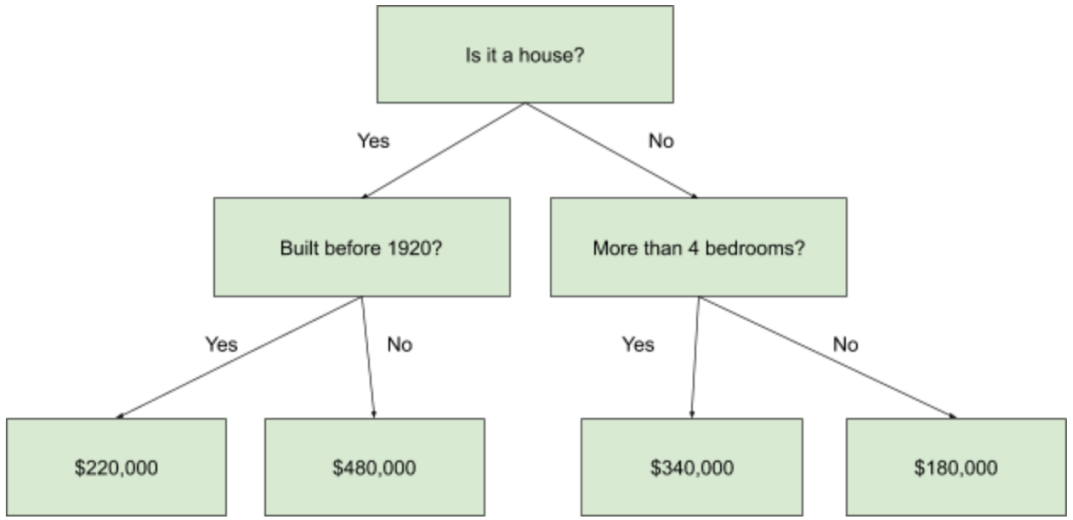
What is machine learning?

Training Data: Database of historical home sales in a county that includes the characteristics of the home and the sales price

Tree 1



Tree 2

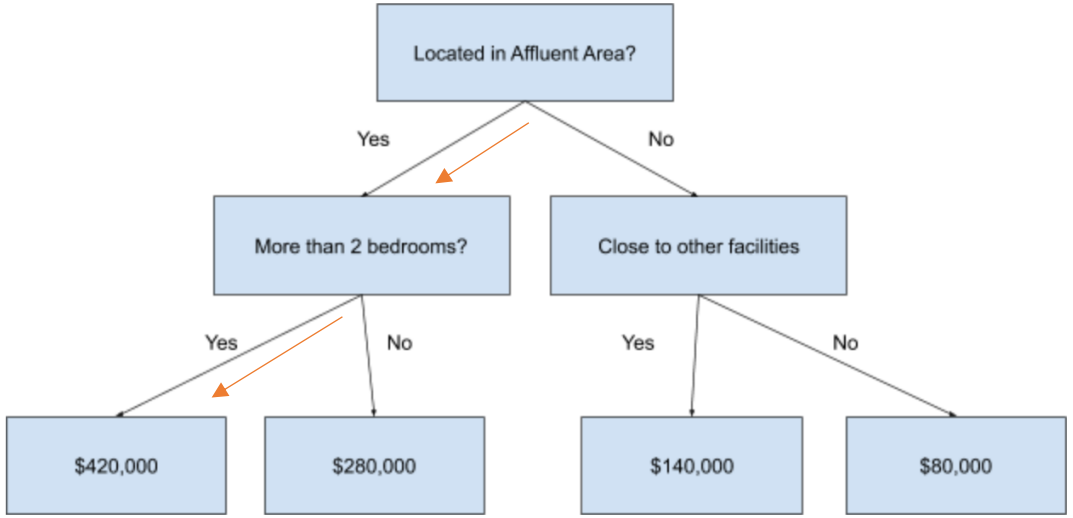


Predicted value is based on the average sales price of the homes in the training dataset that meet the criteria

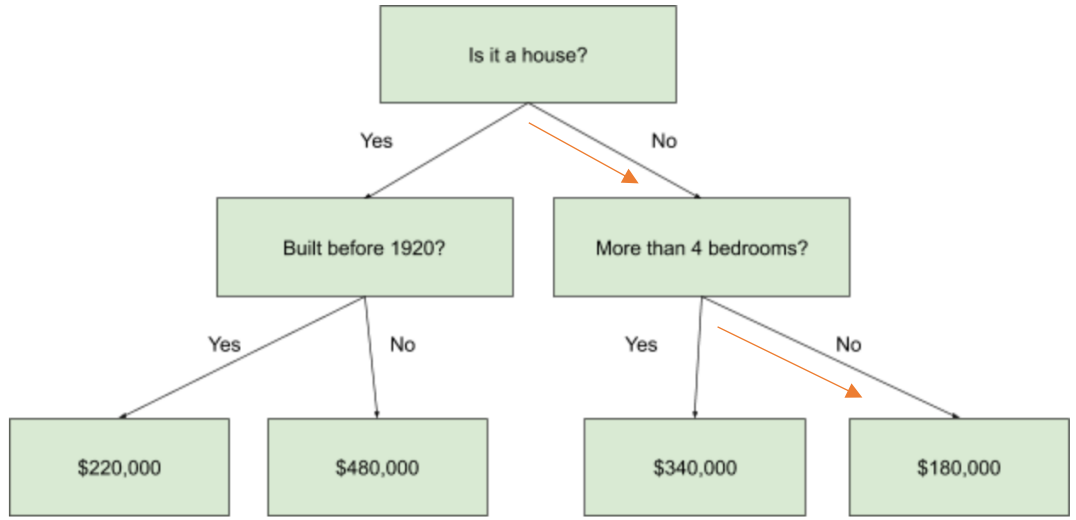
Source: https://www.keboola.com/blog/random-forest-regression

Home for Sale: Condo, affluent area, 3 bedrooms, price is $380,000. Is this a good deal?

Tree 1



Tree 2



Tree 1: Estimated Value: $420,000

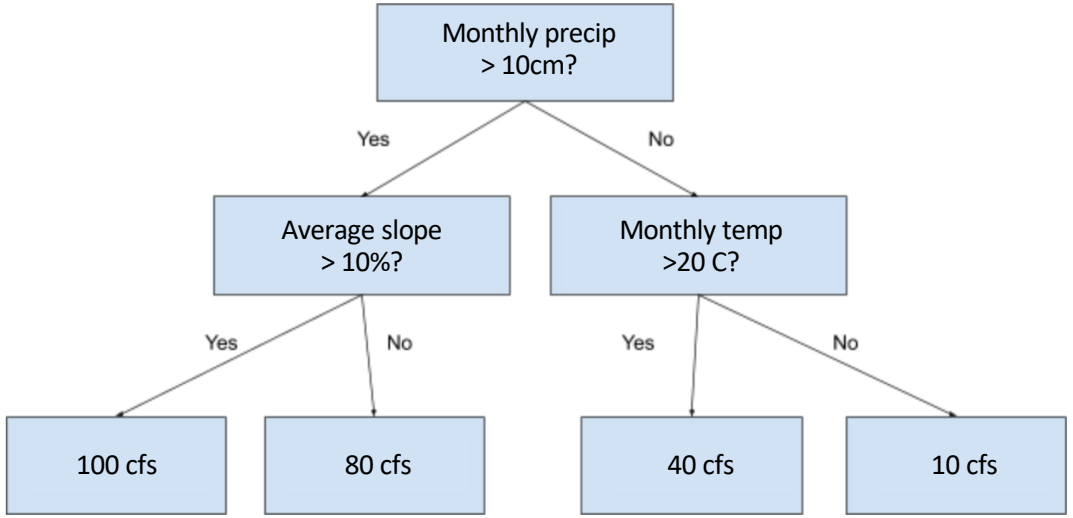
Tree 2: Estimated Value: $180,000

Average Estimated Value: $300,000

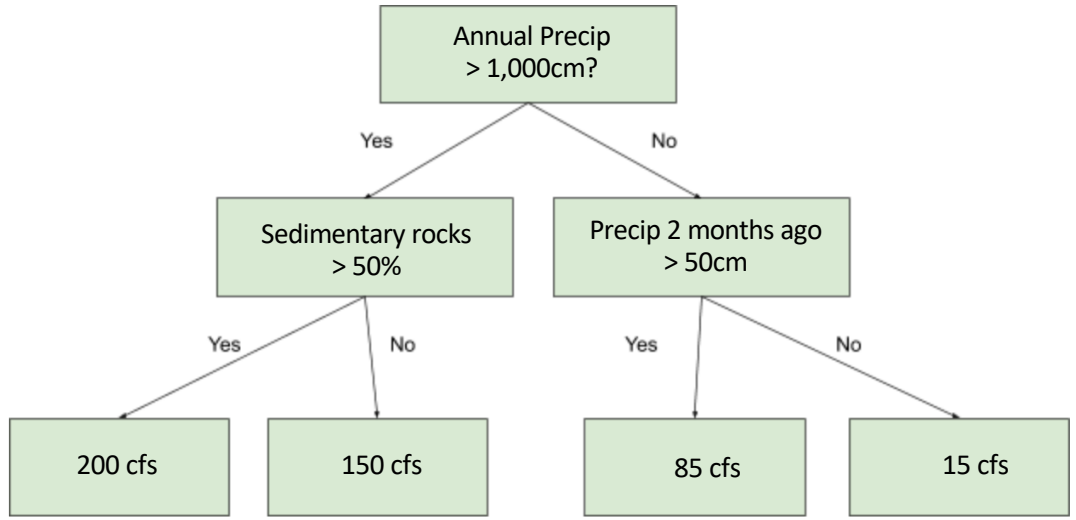
Source: https://www.keboola.com/blog/random-forest-regression

Training Data: Watersheds with minimal human disturbance and a stream gage

Tree 1

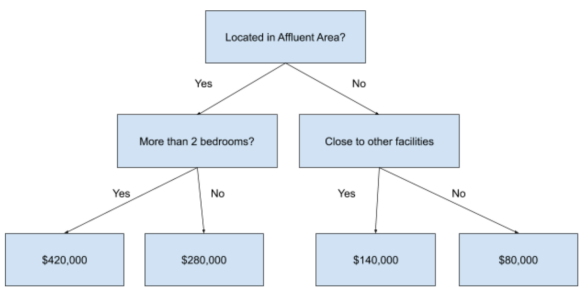


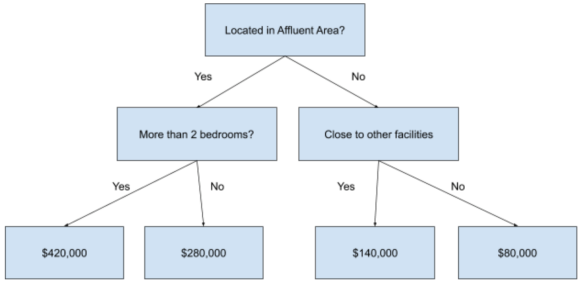
Tree 2

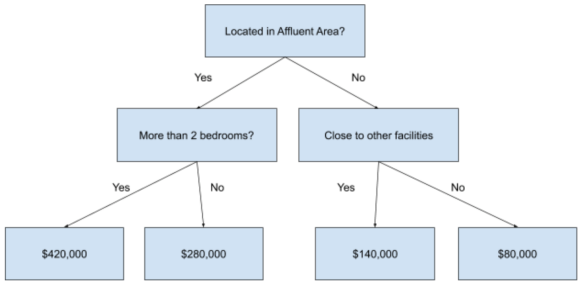


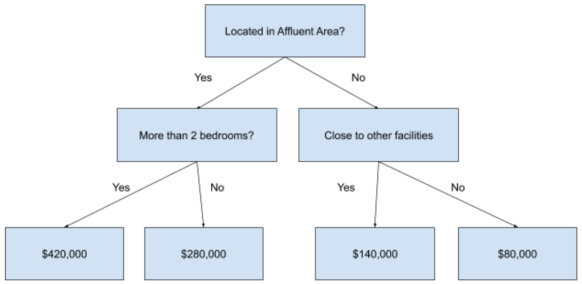
Predicted value is based on the average monthly flow from the stream gages in the training dataset that meet the criteria

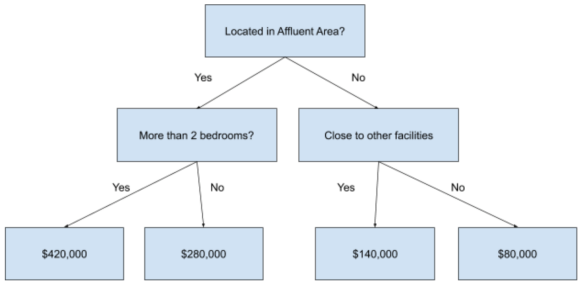
Natural Flows Database: 1,000 trees = 1,000 predictions of flow for every stream segment

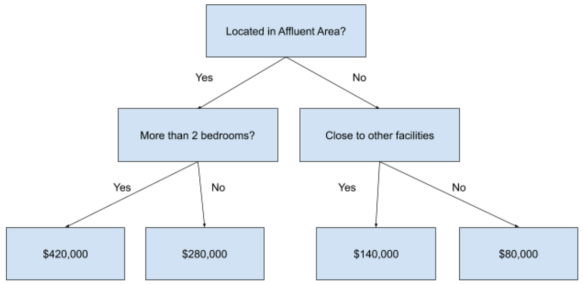












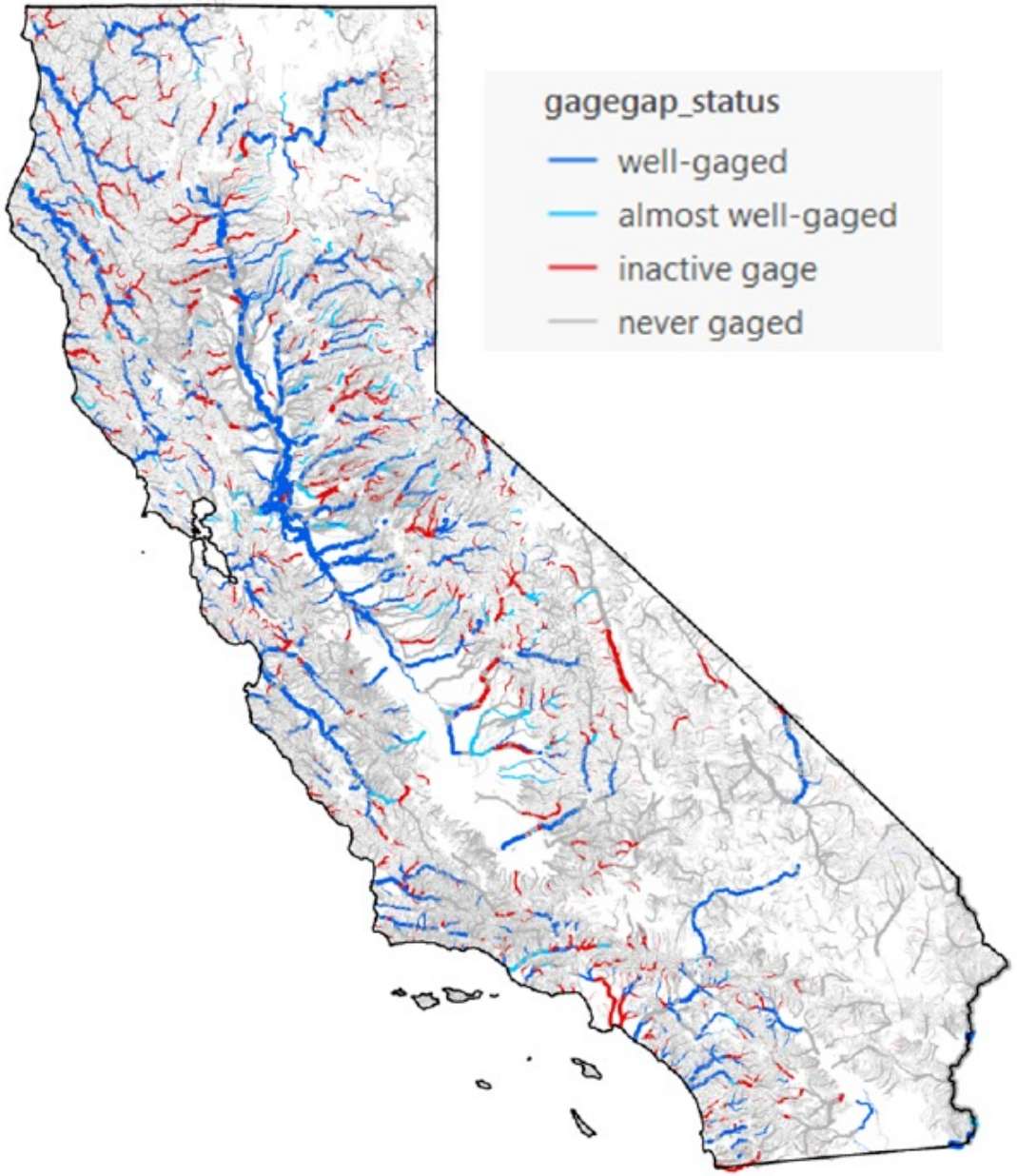
P10: 10th Percentile Prediction

“Estimated”: Average of all the Predictions

P90: 90th Percentile Prediction

Deployment

Apply the models to every stream segment in California



Data Viewer

rivers.codefornature.org

Accepted: 7 November 2017
DOI: 10.1111/fwb.13058

SPECIAL ISSUE

WILEY Freshwater Biology

Patterns and magnitude of flow alteration in California, USA

Julie K. H. Zimmerman[1] | Daren M. Carlisle[2] | Jason T. May[3] | Kirk R. Klausmeyer[4] | Theodore E. Grantham[5] | Larry R. Brown[3] | Jeanette K. Howard[4]

[1]The Nature Conservancy, Sacramento, CA, USA
[2]National Water-Quality Assessment Program, US Geological Survey, Lawrence, KS, USA
[3]California Water Science Center, US Geological Survey, Sacramento, CA, USA
[4]The Nature Conservancy, San Francisco, CA, USA
[5]Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, CA, USA

Correspondence
Julie K. H. Zimmerman, The Nature Conservancy, Sacramento, CA, USA.
Email: julie.zimmerman@tnc.org

Funding information
S.D. Bechtel Jr. Foundation

Abstract

1. Quantifying the natural flow regime is essential for management of water resources and conservation of aquatic ecosystems. Understanding the degree to which anthropogenic activities have altered flows is critical for developing effective conservation strategies. Assessing flow alteration requires estimates of flows expected in the absence of human influence and under current land use and water management.
2. There are several techniques to predict flows in streams and rivers; however, none have been applied to make predictions of natural flow conditions over large regions and time periods. We utilised machine learning statistical models to predict natural monthly flows (natural streamflows without the influence of water management or anthropogenic land use) in California from 1950 to 2015, using time-dependent and fixed watershed variables from reference stream gages. These models were then used to make estimates of mean, maximum and minimum monthly flows in all streams in the state.
3. We compared observed flows measured at 540 stream gages across the state with expected natural flows at the same locations, to quantify the type, frequency and magnitude of flow alteration over the past 20 years (1996–2015). A gage was considered altered if an observed flow metric (monthly mean, annual maximum, annual minimum) fell outside the 80% prediction interval of the modelled flow estimate.
4. We found that 95% of the 540 stream gages in California had at least 1 month of altered flows over the past 20 years, and 11% of gages were frequently altered (over two-thirds of the months recorded had evidence of altered flows). The type of alteration varied across the state with flows being either depleted, inflated or a mix of both at different times of the year. Most altered gages (68%) exhibited both depletion and inflation in monthly flows over the time period. Inflation of monthly mean flows was most prevalent during the summer months, while depletion of monthly flows was evident throughout the year.
5. Type, frequency and magnitude of flow alteration varied by region. Flow depletion was present at >80% of gages in the North Coast and Central Coast, flow inflation was measured at >80% of gages in the South Coast and San Francisco Bay and both depletion and inflation were evident at >80% of gages in the

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2017 The Authors. *Freshwater Biology* Published by John Wiley & Sons Ltd.

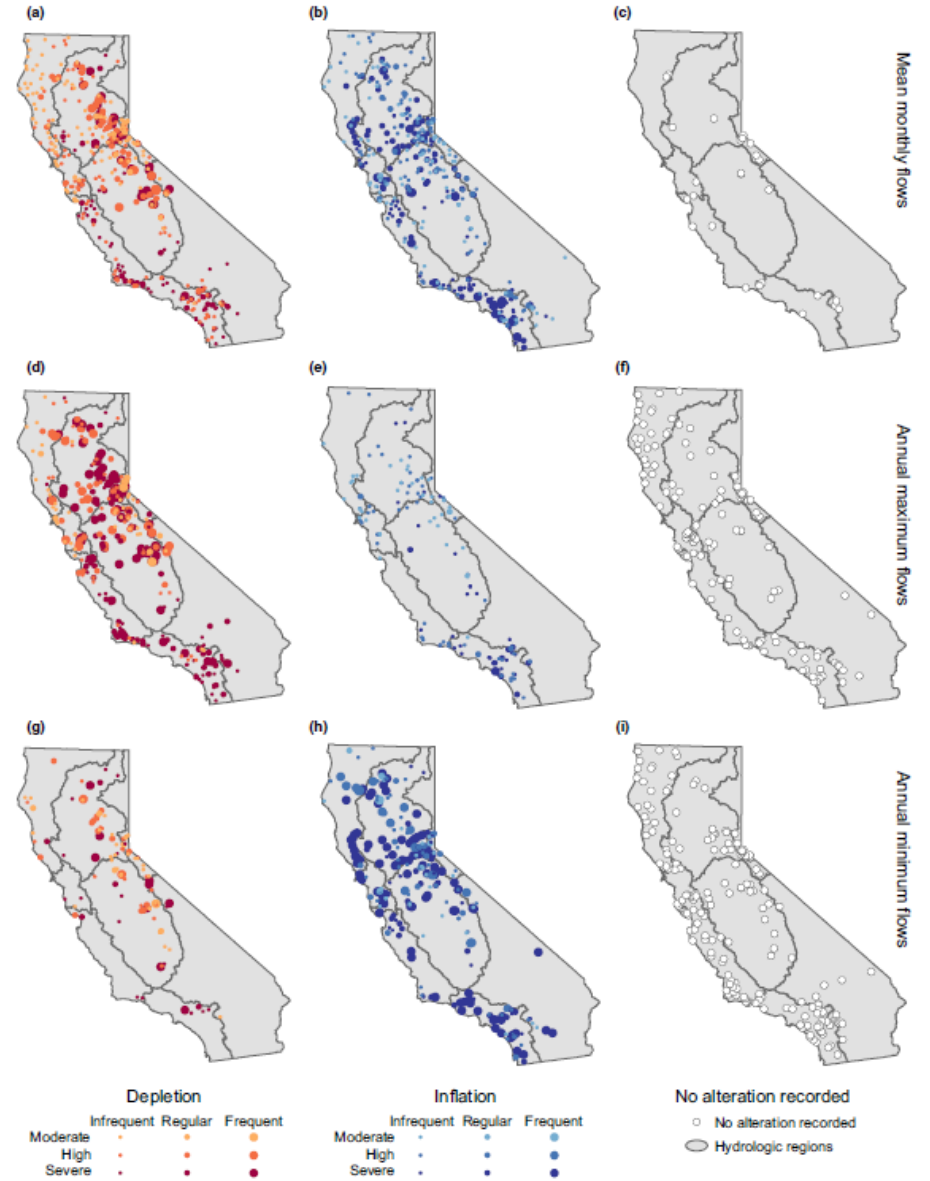


FIGURE 3 Patterns of flow alteration magnitude and frequency for mean monthly (a–c), annual maximum (d–f) and annual minimum (g–i) flows. Alteration frequency is shown by symbol size and magnitude by colour intensity for flow depletion (a,d,g) and inflation (b,e,h). Gage locations with no alteration recorded are also shown (c,f,i)

Instream Flow Regime Criteria on a Watershed Scale

VENTURA RIVER

March 2020

Version 2, updated May 2020

Watershed Criteria Report No. 2020-01

California Department of Fish and Wildlife

Instream Flow Program

Natural Flows

Natural Flows are the streamflows (in cfs) that would be expected with no human influence (data from Zimmerman et al. 2020). Table 1 presents median monthly Natural Flows by month for wet, moderate, and dry water month types for each Ventura River tributary and mainstem reach analyzed in this report. It also presents the drainage area in square miles (mi^2). The numbers next to each stream name correspond to the numbers found on the Ventura River watershed map (Figure 2).

Table 1. Natural Flows.

1) Ventura River 1 225.6 mi^2

	Month Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Natural Flows (cfs)	Wet	97	323	335	166	74	29	10	5	5	5	8	24
	Moderate	19	42	41	30	16	7	2	2	2	2	7	13
	Dry	11	16	14	8	4	2	1	1	1	1	3	7

2) Ventura River 2 214.3 mi^2

	Month Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Natural Flows (cfs)	Wet	103	315	338	166	71	29	10	5	5	5	8	24
	Moderate	18	42	39	29	15	6	2	1	2	2	7	13
	Dry	12	16	13	8	4	2	1	1	1	1	3	7

3) Cañada Larga Creek 19.3 mi^2

	Month Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Natural Flows (cfs)	Wet	5	16	15	9	4	2	1	<1	<1	<1	<1	1
	Moderate	1	3	3	2	1	<1	<1	<1	<1	<1	<1	1
	Dry	1	1	1	<1	<1	<1	<1	<1	<1	<1	<1	1

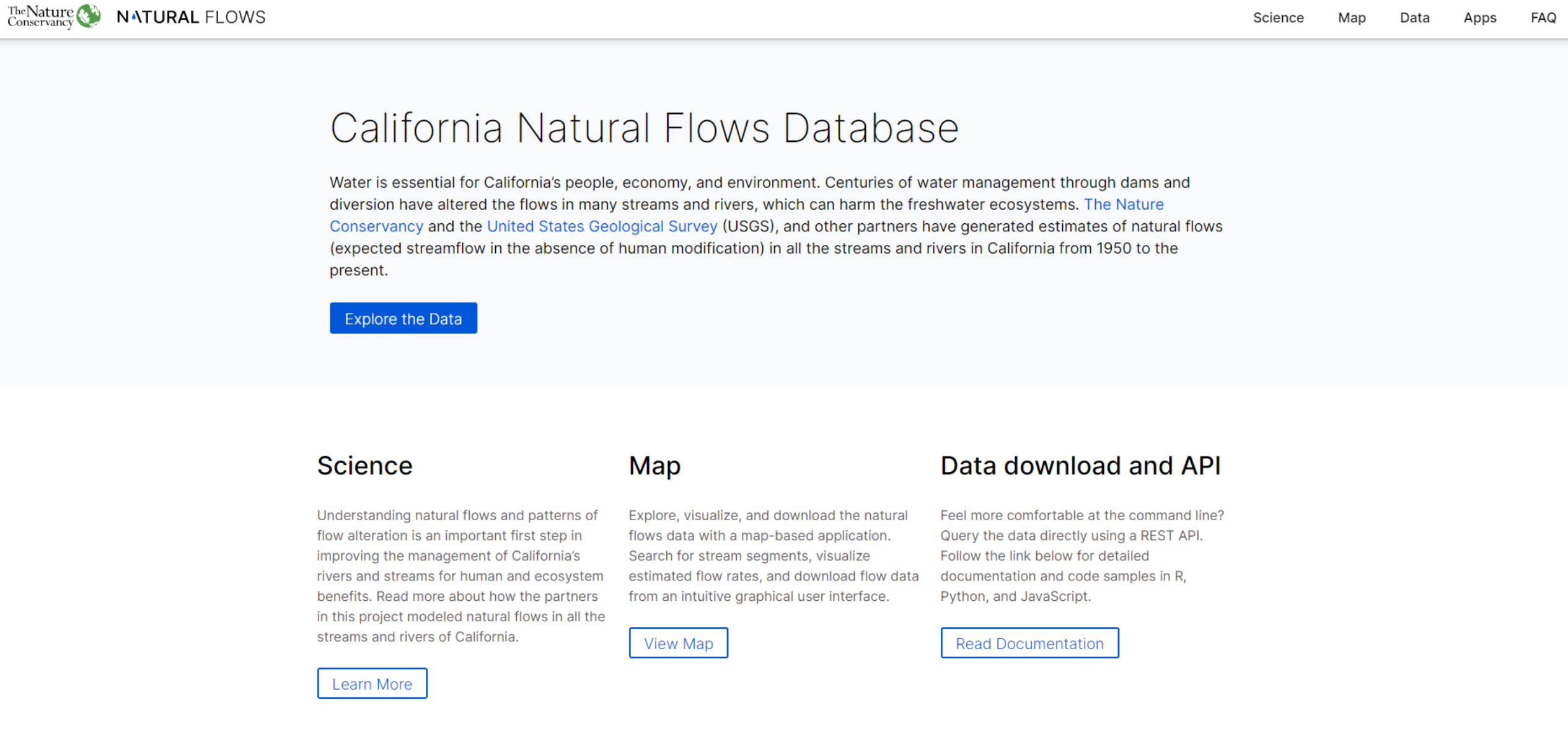
4) Ventura River 3 188.5 mi^2

	Month Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Natural Flows (cfs)	Wet	88	285	316	155	66	28	9	4	5	4	8	22
	Moderate	17	38	36	27	15	6	2	1	1	1	6	11
	Dry	11	15	11	7	4	2	1	1	1	1	3	7

5) Coyote Creek 1 43.5 mi^2

	Month Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Natural Flows (cfs)	Wet	16	49	52	27	11	5	2	1	1	1	1	3
	Moderate	3	8	6	5	3	1	<1	<1	<1	<1	1	2
	Dry	2	3	2	1	1	<1	<1	<1	<1	<1	1	1

6) Ventura River 4 141.6 mi^2

	Month Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Natural Flows (cfs)	Wet	72	218	268	143	57	24	9	4	4	4	6	20
	Moderate	13	30	31	23	12	5	2	1	1	1	5	8
	Dry	9	12	10	6	3	2	1	<1	1	1	2	5

https://rivers.codefornature.org