

STAT 215A Fall 2019

Week 8b

Tiffany Tang

10/17/19

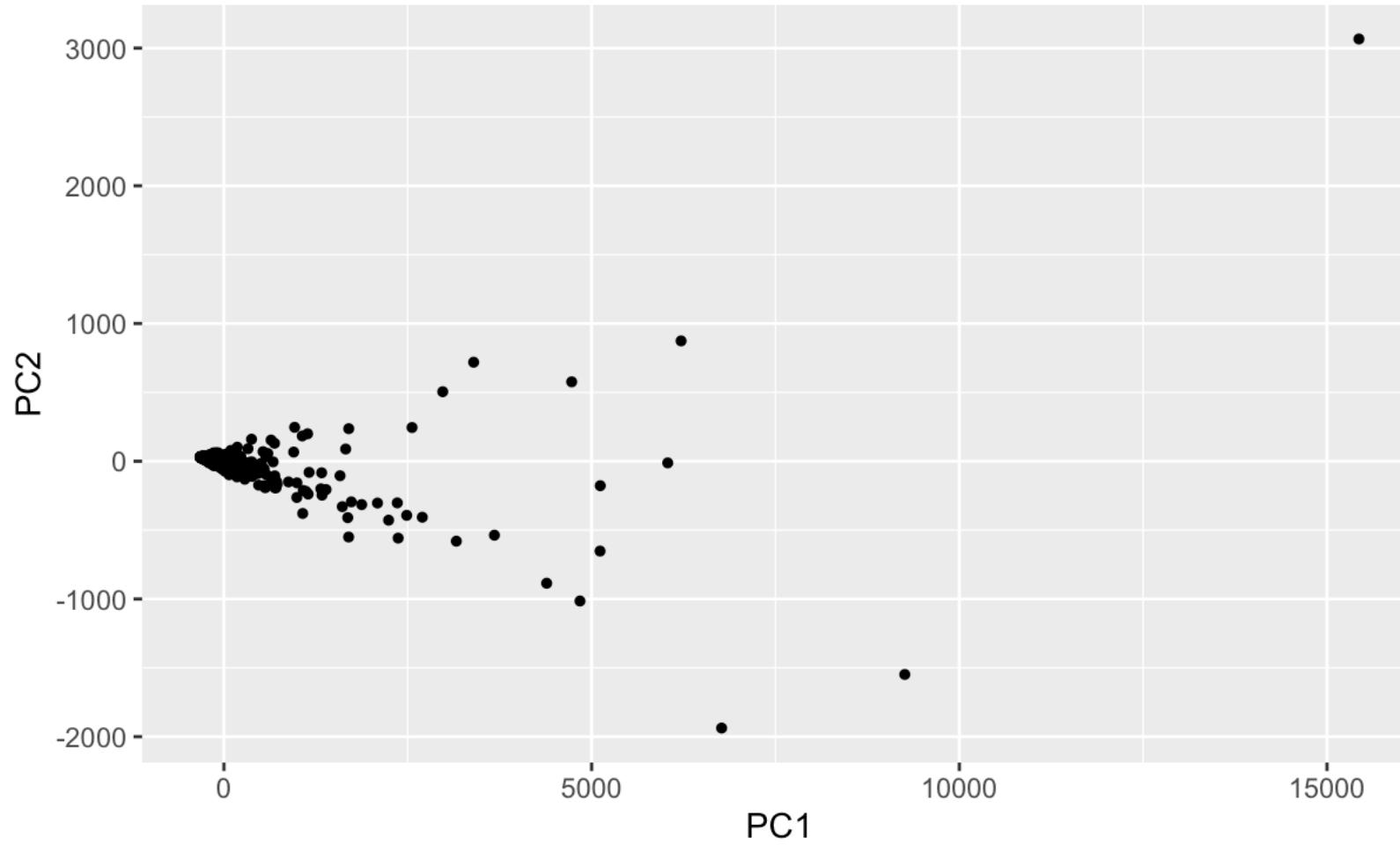
Lab 3 – Stability of K-means + Computability

- ▶ Sign up for an SCF account if you haven't already
- ▶ Don't wait until the last minute to do the lab
 - ▶ SCF could be busy and it takes time for your code to run
- ▶ No need to do PCA; just apply k-means to raw lingBinary data
- ▶ You can do better than the Figure 3 in Ben-Hur
- ▶ Either manually copy over the data/ folder to SCF or push it to GitHub and then remove it
- ▶ While the writeup is shorter than usual, there will still be a writing component of the grade
- ▶ Take a look at [Google R Style Guide](#) and Part I Analyses Section of the [Tidyverse Style Guide](#)
- ▶ If you are using the SCF JupyterHub, be sure to “Stop Server” when you are done

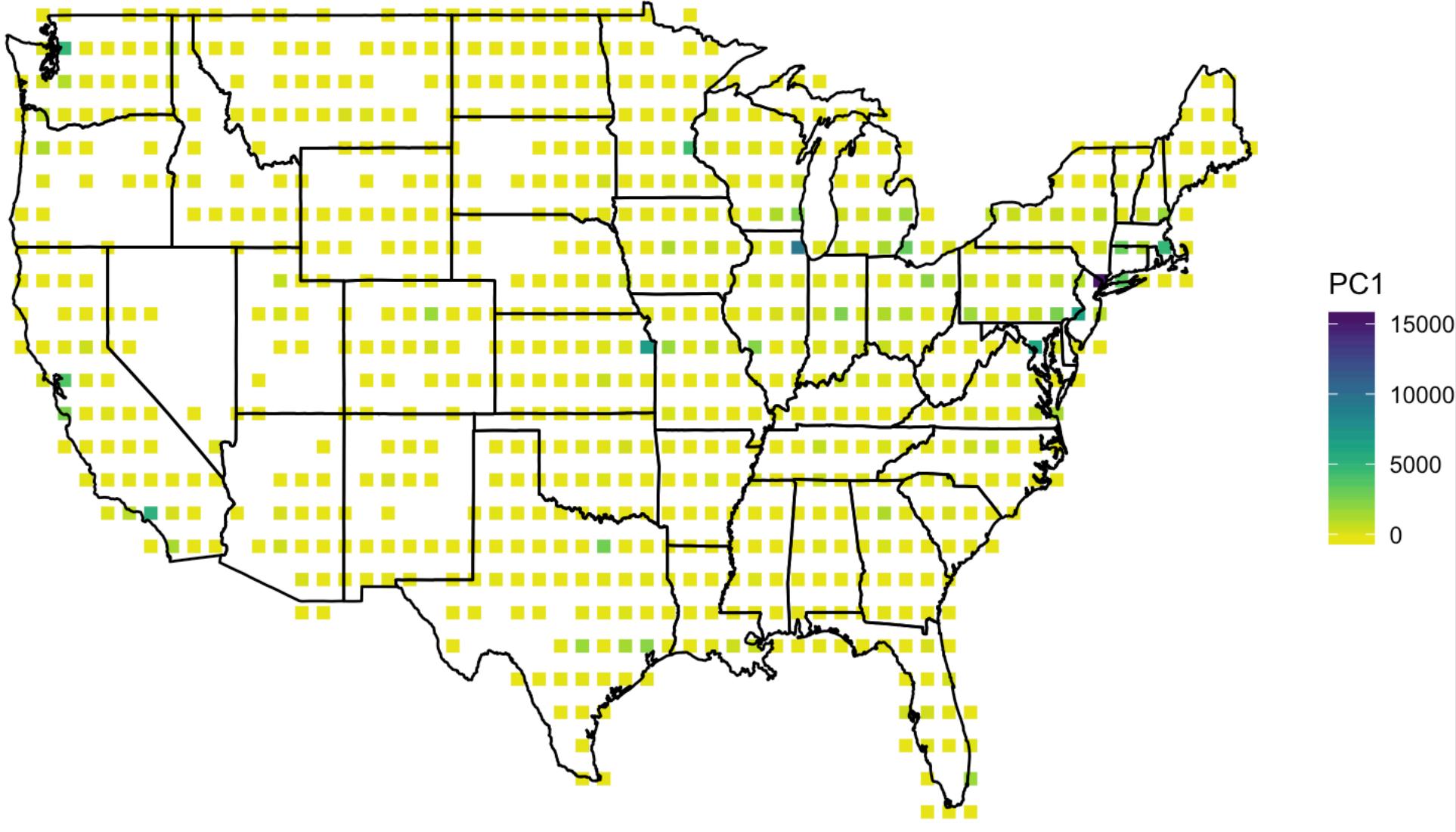
Lab 2 – Linguistics Data

- ▶ I've enjoyed reading your reports thus far! Some interesting analyses
- ▶ General tip when plotting: try not to use too many colors in discrete color scheme
 - ▶ ~8 is a maximum but depending on the color scheme, even 8 different colors can make it hard to read the plot
- ▶ Even though unanswered questions are given to you as 0s, these should be treated as NAs
- ▶ PCA = **P**rincipal Components Analysis (not principle)
- ▶ Be careful when aggregating data

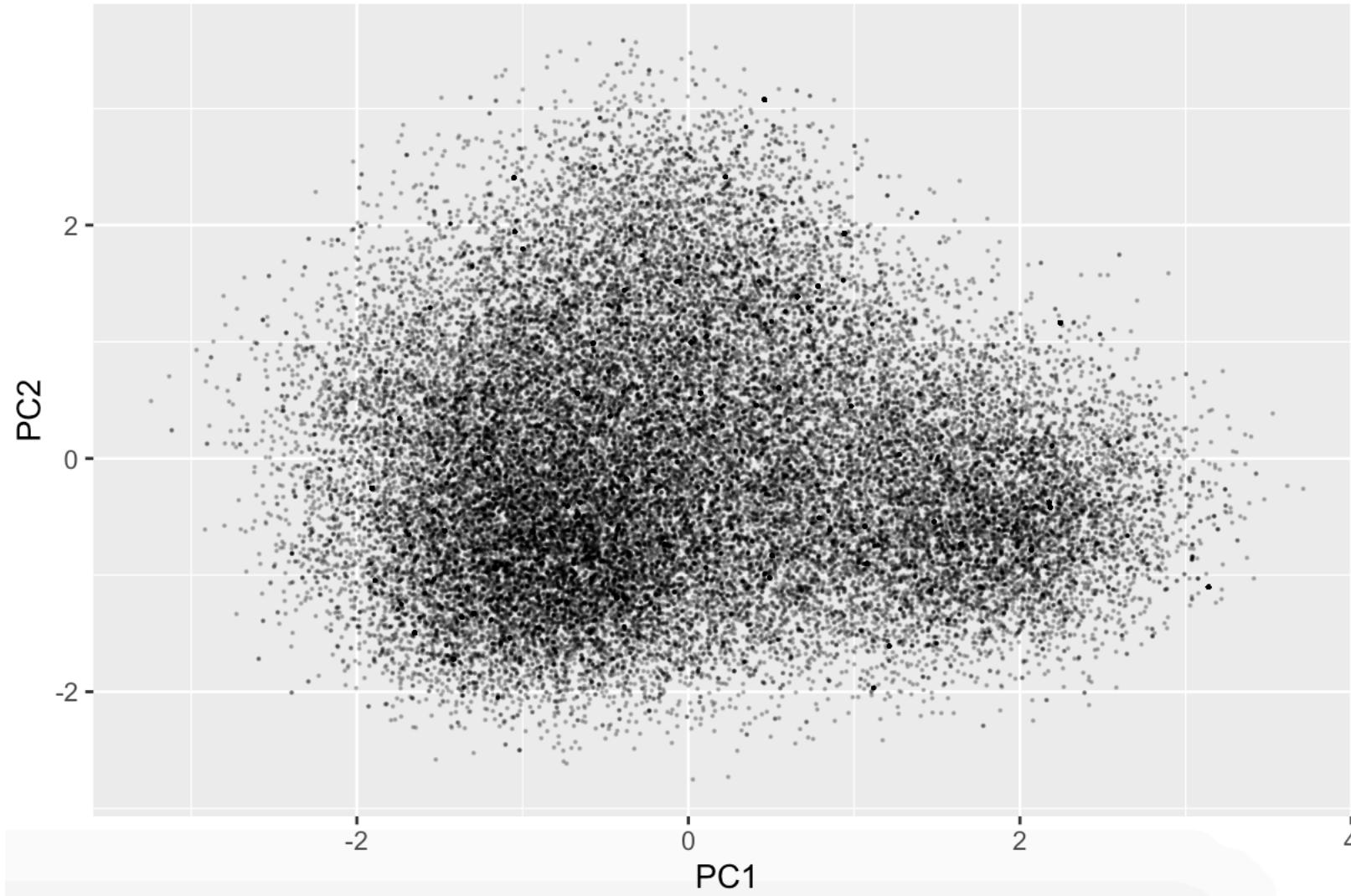
PCA on lingLocation



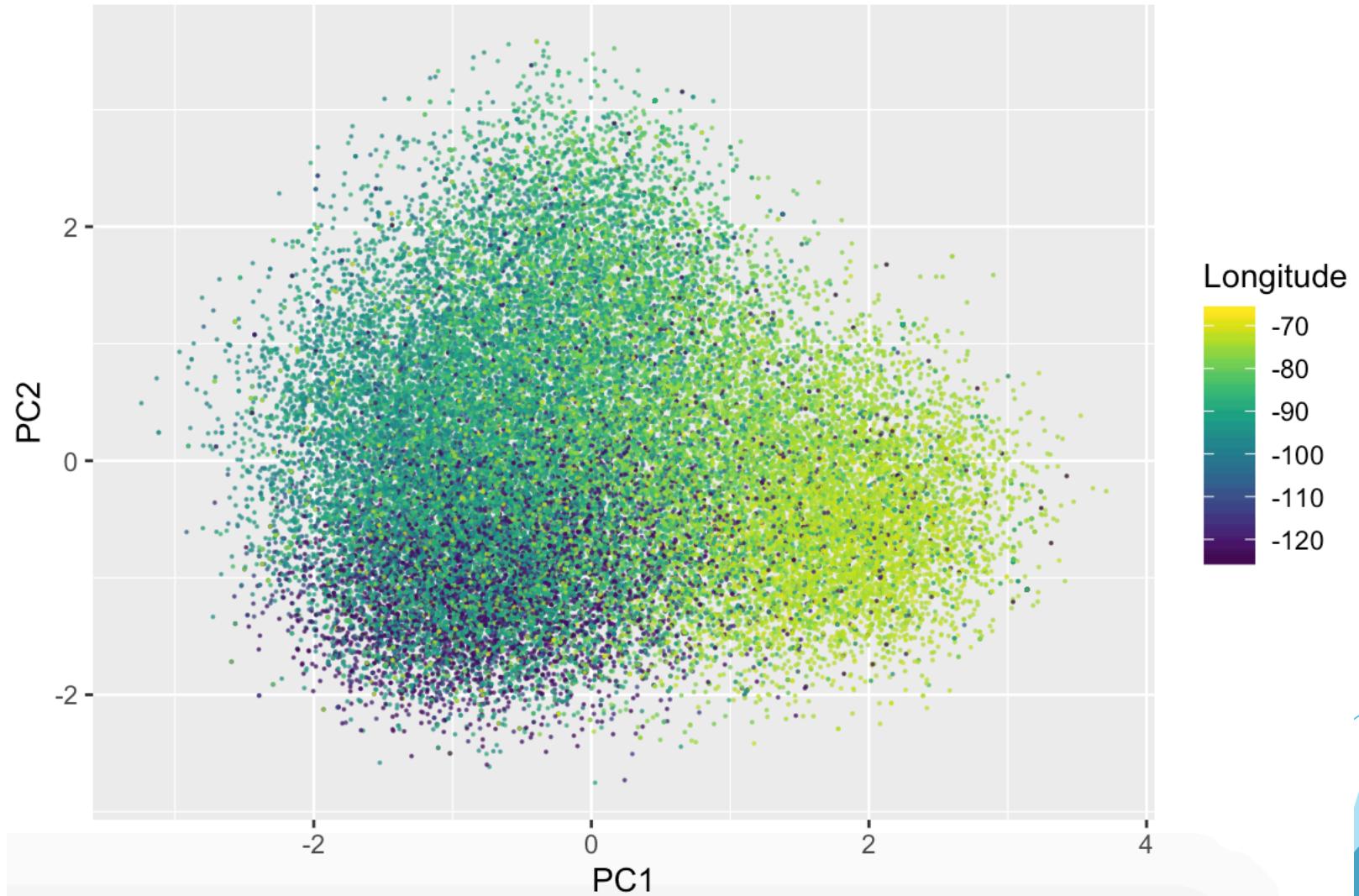
PCA on lingLocation



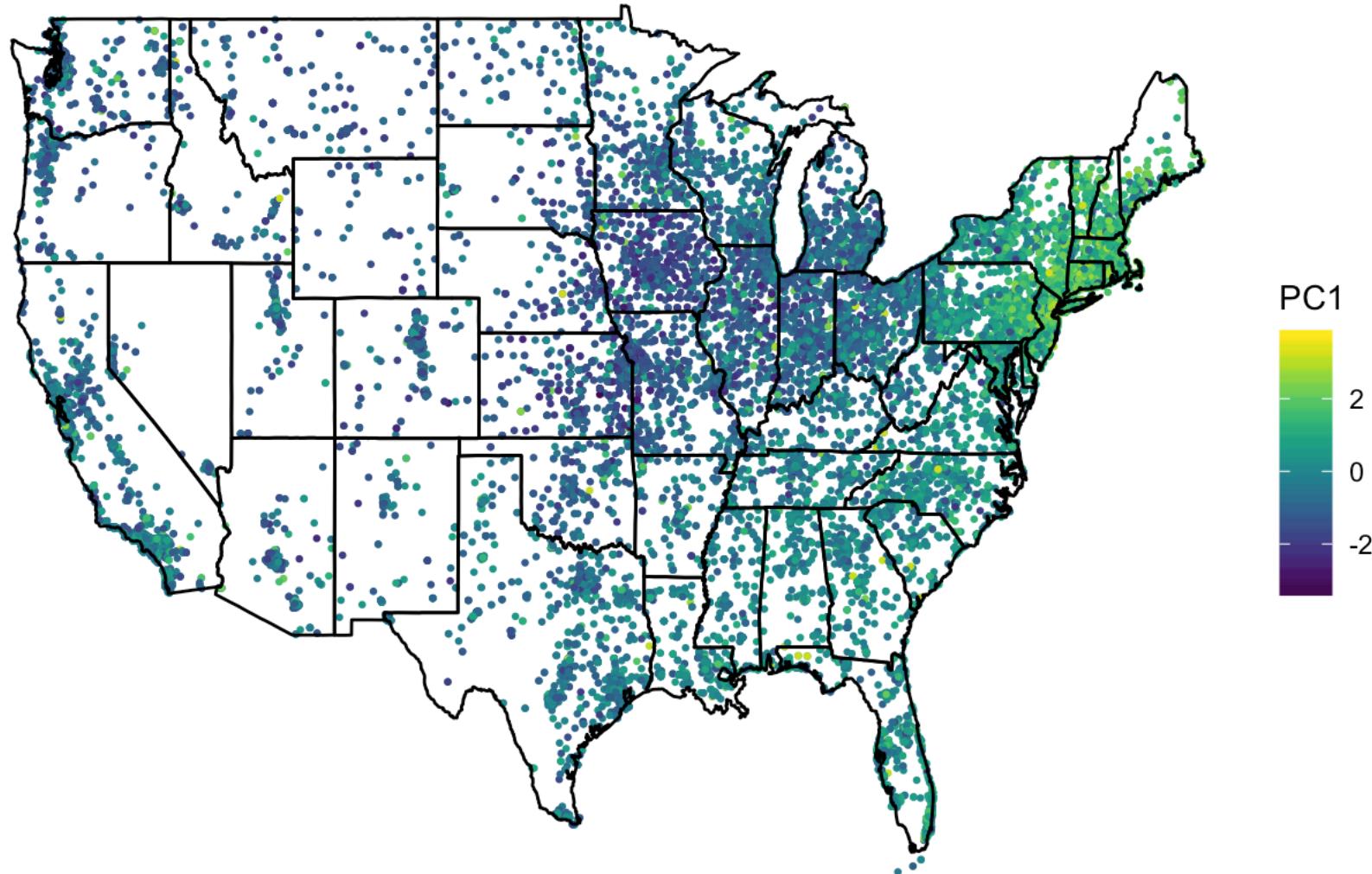
PCA on lingData



PCA on lingData



PCA on lingData



Plan for Today:

- ▶ UCERF3 Earthquake report
- ▶ Efron (1979) Bootstrap paper

30th Anniversary of Loma Prieta



EARTHQUAKE EXTRA

San Francisco Chronicle

The Largest Daily Circulation in Northern California

125th Year No. 237

THURSDAY, OCTOBER 19, 1989

415-777-1111 25 CENTS

Photo by AP Wirephoto

DEVASTATING REPORTS FROM BIG QUAKE AREA

Disaster Declared in 7 Counties

By Randy Shilts
and David Ward
Chronicle Staff Writers

A battered but determined Bay Area faced with mounting casualties, rising death tolls, thousands of tons of shattered Oakland freeway. Thirty bodies were recovered from the scene, but officials predicted hundreds.

The death toll remained uncertain last night because most victims were still buried under thousands of tons of shattered Oakland freeway. Thirty bodies were recovered from the scene, but officials predicted hundreds.

At least 1,400 were injured by the quake's severe jolt in the region since the great San Francisco earthquake. Death toll estimates climbed to more than 100, with the deviation in the number alone estimated at 52.

Geologists have already given the 30-second quake a name: "The Great Quake." It hit the seven hardest-hit counties, stretching from Monterey and San Benito in the south to Santa Clara and San Francisco in the north. The declaration opened the way for delivery of billions of dollars in federal relief funds.

An expression of sympathy and offers of help poured in from around the world. President Bush declared a state of emergency in the seven hardest-hit counties, stretching from Monterey and San Benito in the south to Santa Clara and San Francisco in the north. The declaration opened the way for delivery of billions of dollars in federal relief funds.

In spite of the outrage, The Chronicle was published despite a second day of complete power outages at the newspaper's headquarters at Fifth and Mission streets. The disruption of electrical service to the area took down the paper's computers and main printing facility.

"We will take every step and measure to get back to the Bay Area in its hour of need," Bush said, adding that he may tour the area Saturday.

From Santa Cruz to Oakland and San Francisco, emergency crews searched through mounds of twisted metal and concrete in desperate attempts to find survivors.

Even though yesterday's quake was not the biggest to be expected, many geologists were surprised at the extent of the damage up to 50 miles away. "It was an unexpected earthquake," said Jerry Ulrich, a geologist with the U.S. Geological Survey. "I had no idea it would knock down freeways."

While the underground pressure that caused the entire San Andreas fault to move along the pattern lessened, smaller strands are still popping up along other faults.

Early scientists worked around the clock yesterday to assess the exact damage. As the quake, geologists emphasized that this is not the "big one" that they have been fearing for years to come to face with the reality of life in the most fragile of cities ever happened.

The price we pay for being San Francisco is the ever-present threat of instant disaster.

"This is just how the planet works around here," said William Ellsworth at the United States Geological Survey's research center in Menlo Park. Geologists say that the quake registered 7.1 on the Richter scale, and yet there was little damage.

"The government geologists gave the Loma Prieta a 100 percent chance of hitting the San Francisco area again," said Ellsworth.

Working without electricity, the staff of the Chronicle worked through the night to put the paper together, and got it delivered, under very primitive and difficult circumstances, to newsstands and libraries of them - as well as its gratitude to the readers for their continued understanding.

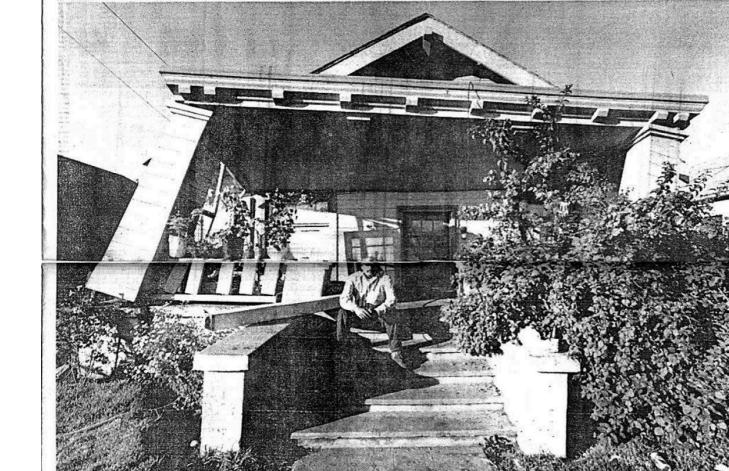
Crews at the expected to find the remains of more bodies in as many as 250 vehicles trapped beneath the area of the highway. They said about 30 bodies had been recovered, and could be found as the search would probably climb dra-

maticaly.

On Tuesday, state officials put the death tally at more than 200, but yesterday they called it count of 100 bodies found. They said about 30 bodies had been recovered, and could be found as the search would probably climb dra-

matically.

Because so many deaths are



Guillermo Vega of Watsonville sat on the front steps of his house, which was knocked askew by Tuesday's quake.



It Hit Where Experts Expected

By Charles Petit
Chronicle Science Writer

Days of Our Years

Government scientists say yesterday's quake hit exactly where they had considered one of the Bay Area's most dangerous stretches of the Western continent.

"It was an expected earthquake," said Jerry Ulrich, a geologist with the U.S. Geological Survey. "But I wouldn't have thought it would knock down freeways."

Even though yesterday's quake was not the biggest to be expected, many geologists were surprised at the extent of the damage up to 50 miles away. "It was an unexpected earthquake," said Jerry Ulrich, a geologist with the U.S. Geological Survey. "But I wouldn't have thought it would knock down freeways."

While scientists worked around the clock yesterday to assess the exact damage. As the quake, geologists emphasized that this is not the "big one" that they have been fearing for years to come to face with the reality of life in the most fragile of cities ever happened.

The price we pay for being San Francisco is the ever-present threat of instant disaster.

"This is just how the planet works around here," said William Ellsworth at the United States Geological Survey's research center in Menlo Park. Geologists say that the quake registered 7.1 on the Richter scale, and yet there was little damage.

"The government geologists gave the Loma Prieta a 100 percent chance of hitting the San Francisco area again," said Ellsworth.

Working without electricity, the staff of the Chronicle worked through the night to put the paper together, and got it delivered, under very primitive and difficult circumstances, to newsstands and libraries of them - as well as its gratitude to the readers for their continued understanding.

Crews at the expected to find the remains of more

bodies in as many as 250 vehicles trapped beneath the area of the highway. They said about 30 bodies had been recovered, and could be found as the search would probably climb dra-

matically.

On Tuesday, state officials put the death tally at more than 200, but yesterday they called it count of 100 bodies found. They said about 30 bodies had been recovered, and could be found as the search would probably climb dra-

matically.

Because so many deaths are

SEE BUSINESS, PAGE A 12

World Series May Return Tomorrow

at Candlestick

SEE SPORTS, PAGE A 13

Many Tourists Fled San Francisco in Fear

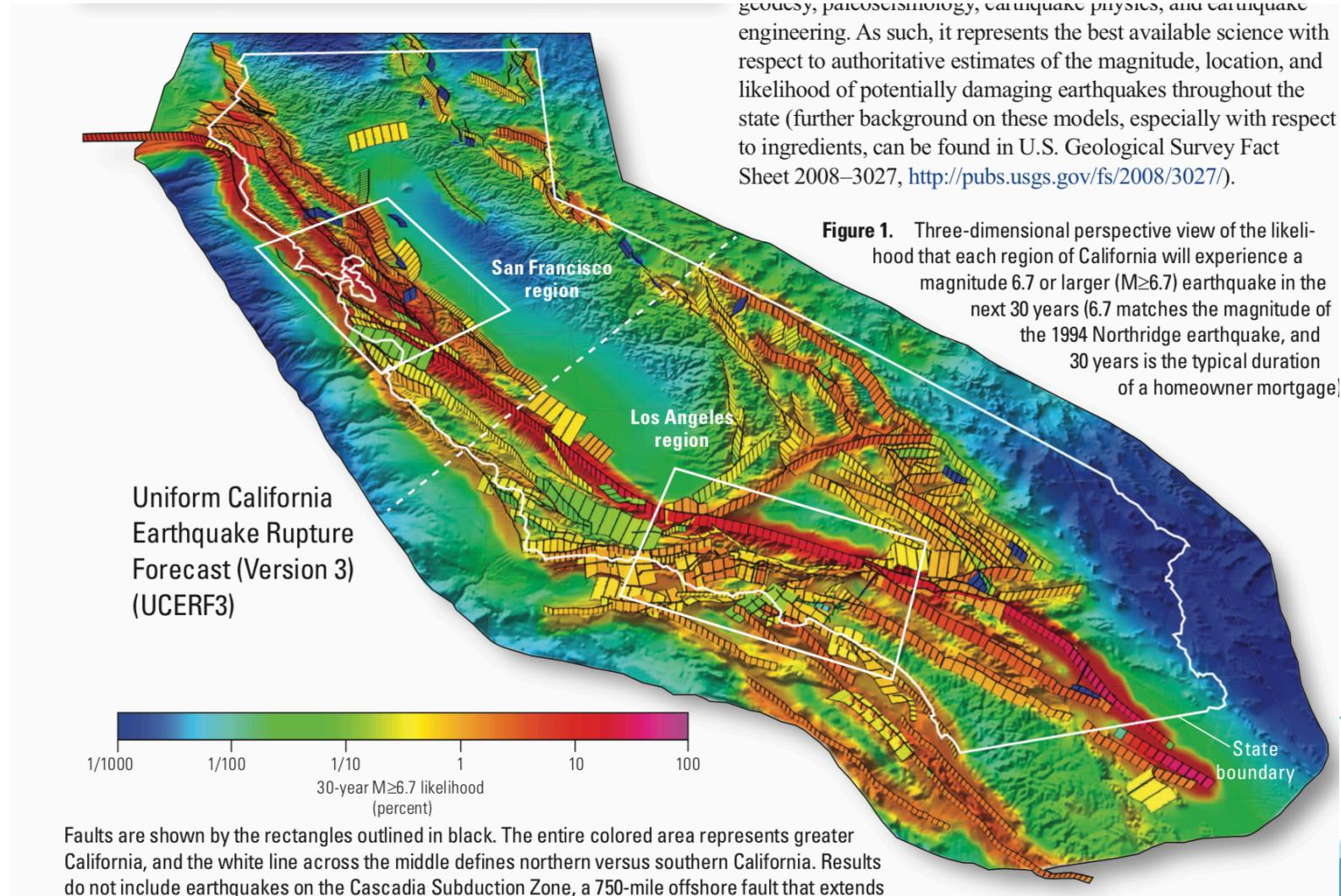
SEE PEOPLE, PAGE A 14

Hurricane Reaches East Germany

SEE WORLD/NATIONAL NEWS, PAGE A 15

SEE WORLD, PAGE A 16 Col. 1

UCERF3: A New Earthquake Forecast for California's Complex Fault System



UCERF3: A New Earthquake Forecast for California's Complex Fault System

1. What is new about this updated earthquake model compared to previous models?
2. What domain knowledge was brought into the modeling aspects of UCERF3?
3. In what ways is the most recent model still not ideal/perfect? What are some challenges that arise when trying to incorporate domain knowledge into models?
4. What are things that the USGS does well with their UCERF3 work that can be transferred to the data science problems that you imagine working on in the future?

Efron's (1979) Bootstrap Paper

- ▶ **What is the bootstrap?**
 - ▶ We've been using bootstrap as a means of perturbing the data
 - ▶ Usually used as a computational approach for estimating SEs, bias, confidence intervals, hypothesis testing
- ▶ Today and in this paper, focus on estimation
- ▶ **Motivating Example:** Given n iid data points $X_i \sim Bern(p)$, want $\text{Var}(\bar{X})$
 - ▶ From the theory, $\text{Var}(\bar{X}) = \frac{p(1-p)}{n}$
 - ▶ But what if we didn't know that formula?
 - ▶ Use the bootstrap procedure

Efron's (1979) Bootstrap Paper

1. Why does the bootstrap sample have to be of size n , that is, the sample size of the original data?
2. Can we use the bootstrap with a really large number of bootstraps B to fix small sample size problems?
3. What would you do if your data samples are not independent?