

## PROCESS BOOK OPEN LINK

Names: Nathaniel Berkman ([nberkman@college.harvard.edu](mailto:nberkman@college.harvard.edu)), Catherine Mignone ([catherinemignone@college.harvard.edu](mailto:catherinemignone@college.harvard.edu)), Matteo Wakeman ([matteowakeman@college.harvard.edu](mailto:matteowakeman@college.harvard.edu))

## Milestone 11:

1. **Complete data pipeline:** We have finished cleaning our data, as you can see in the dataset in the zip file.
2. **Functional D3 visualizations:** The D3 visualizations are functional and integrated in the webpage layout and structure.
3. **Website structure:** We added in keyboard shortcuts for different tabs and also a click-through feature. The maps are interactive.
4. **Clear storytelling:** We think that our narrative of election dominance to loss comes across clearly.

We also finished major bug fixes and incorporated the layout from

## Milestone 10:

1. **Team member documentation:** Catherine Mignone, Matteo Wakeman, Nathaniel Berkman
2. **Complete data pipeline:** We included our geoson file.
3. **Functional D3 visualizations:** Included in the website!

We made large changes to the prototype between Milestones 9 and 10, and are very proud of the current prototype. We had to do some data-cleanup that we only noticed once we engaged with more advanced visualizations. The website is now substantially more interactive. One tactical change that we made was to make our visualizations and analysis more robust, we included the ability to download the geojson file that we worked with from the website so that any viewer could test our hypotheses. We are looking forward to hearing your feedback on this prototype.

## Milestone 9:

### Overview:

For milestone 9, we started compiling a functional visualization prototype of our data story, "Unity in Central Namibia: The Path to Power". With an emphasis on vote fragmentation, geographic clustering, and turnout disparities, the project explores why the SWAPO party continues to hold political sway in spite of quantifiable declines in voter support.

**Worked on Visualizations:** Nathaniel Berkman, Catherine Mignone, Matteo Wakeman

**Data:**

Our main dataset, `namibia.geojson`, contains the official constituency boundaries for Namibia. Each shape on the map represents one of the 121 constituencies, with information about the region name and coordinates.

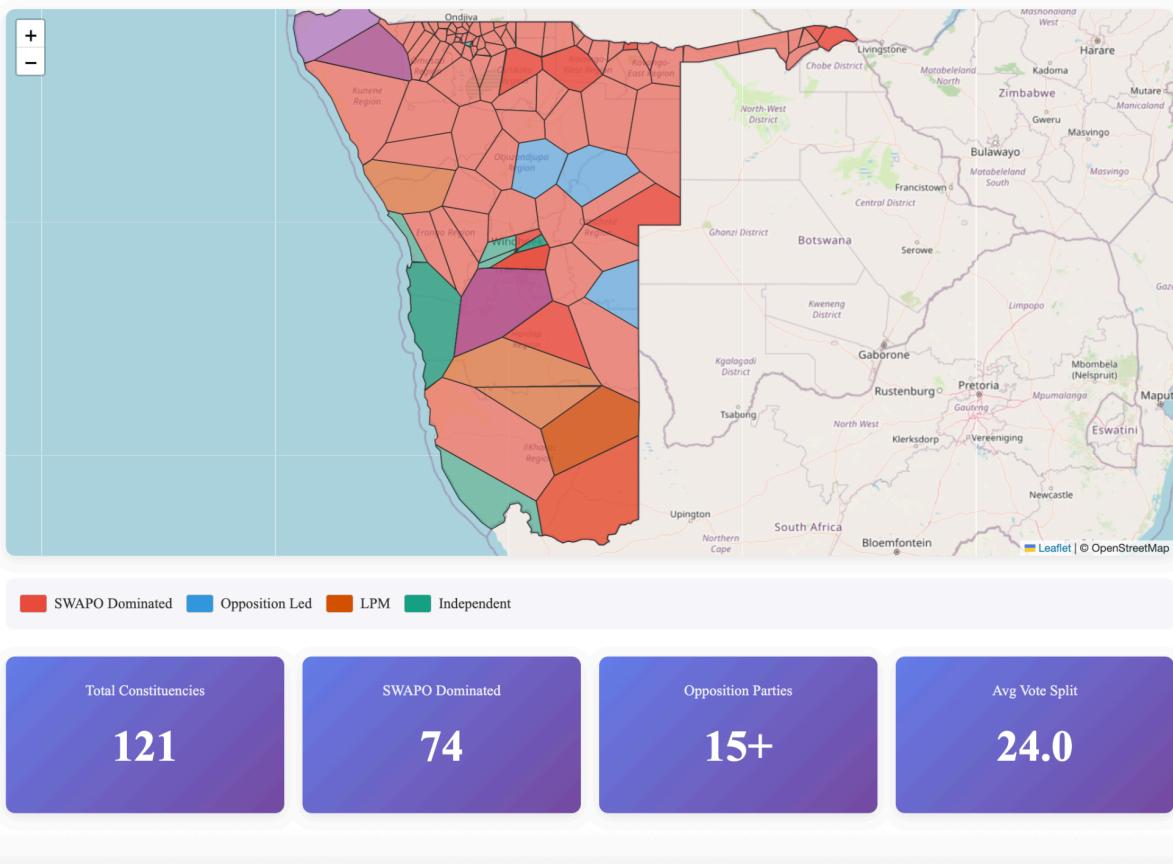
This dataset allowed us to show results such as:

- Which party won each area
  - How vote share changed between years
- Where elections were close (small margins of victory)
- How turnout increased or dropped
  - How many parties competed in each constituency

**Images of the Visualizations on the Website:**

## The Current Reality: Fragmented Opposition

The 2019 Namibian elections revealed a critical challenge for opposition parties: vote splitting across multiple parties has allowed SWAPO to maintain dominance, even in constituencies where combined opposition votes could have prevailed.



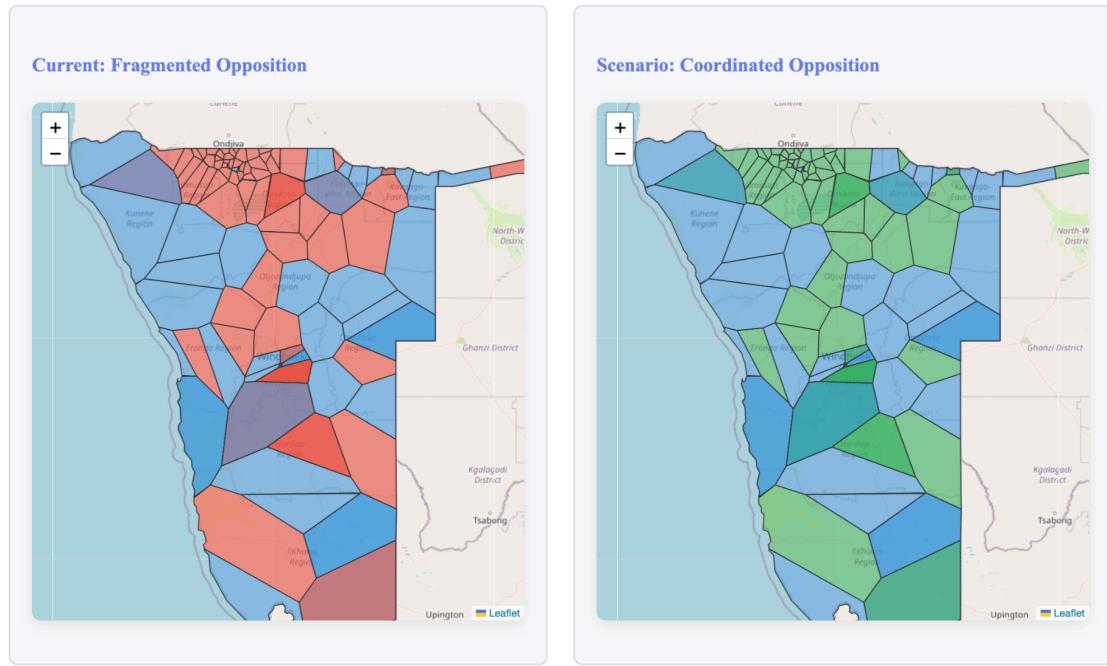
### Fragmented Opposition Map

A color-coded map of all 121 constituencies showing which party won in 2019.

- Red = SWAPO, Blue = Opposition, Orange = LPM, Green = Independent
- Shows how split opposition votes helped SWAPO stay in power
- Summary boxes below show total constituencies, SWAPO-held seats, number of opposition parties, and the average vote split

## What If: The Power of Coordination

Analysis of the top 2 opposition parties shows the dramatic impact that strategic coordination could achieve.



### Impact with Coordination:

- **74 constituencies**  
would flip to opposition control
- **0 more seats**  
would become genuinely competitive (within 5% margin)
- Potential to reduce SWAPO's parliamentary majority significantly

## Coordinated Opposition Scenario

Two side-by-side maps: one shows the real 2019 results, and the other shows what would happen if the top two opposition parties joined forces.

- Around 74 constituencies would flip under coordination

## Top Target Constituencies

### 1. Aranos

Hardap



Winnability: 216/100

### 2. Arandis

Erlongo



Winnability: 211/100

### 3. Opuwo Urban

Kunene



Winnability: 206/100

### 4. Rehoboth Rural

Hardap



Winnability: 205/100

### 5. John Pandeni

Khomas



Winnability: 200/100

### 6. Mariental Rural

Hardap



Winnability: 199/100

### 7. Mariental Urban

Hardap



Winnability: 198/100

### 8. Katutura East

Khomas



Winnability: 192/100

### 9. Karibib

Erlongo



Winnability: 189/100

### 10. !Nam!Nûs

!Karas



Winnability: 188/100

### 11. Omaruru

Erlongo



Winnability: 187/100

### 12. Rehoboth Urban East

Hardap



Winnability: 185/100

### 13. Gobabis

Omahaheke



Winnability: 177/100

### 14. Omatako

Otjozondjupa



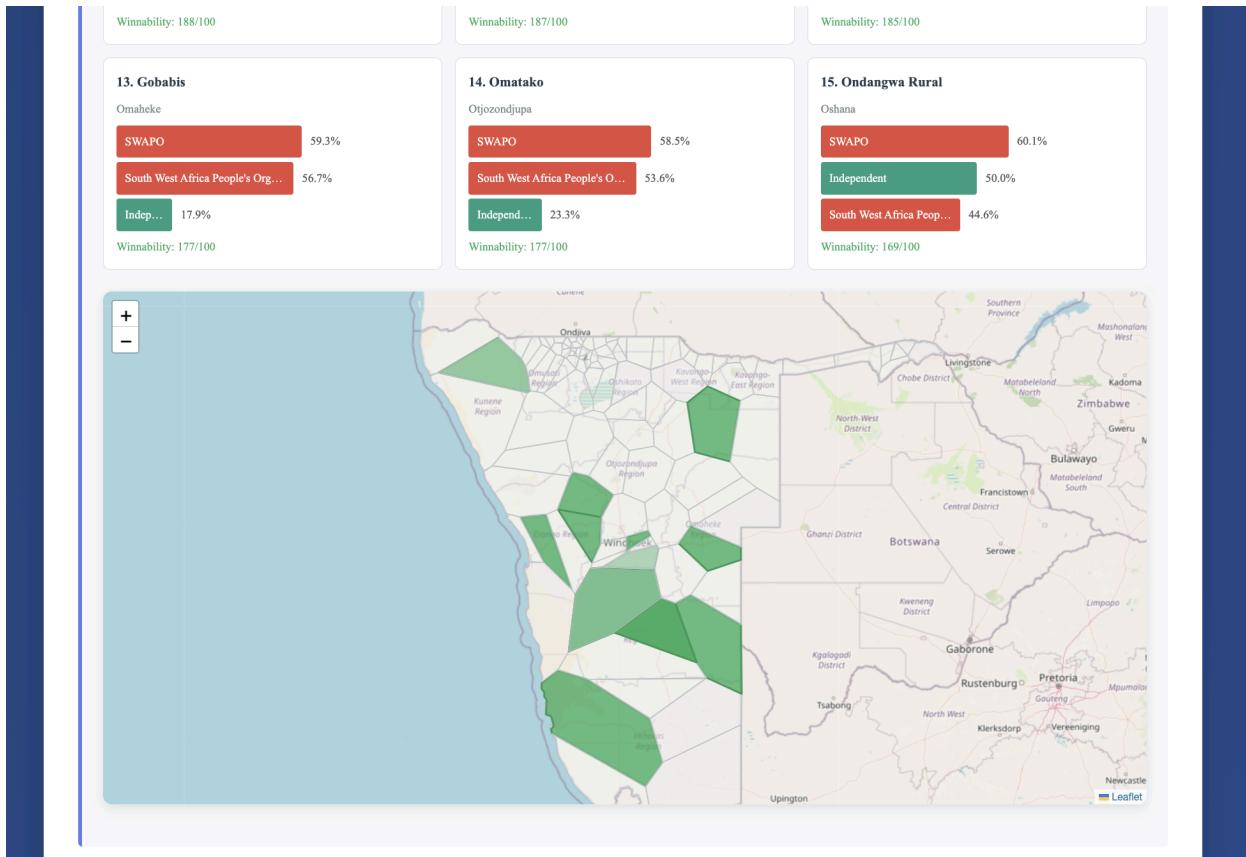
Winnability: 177/100

### 15. Ondangwa Rural

Oshana



Winnability: 169/100



## Target Constituency Dashboard

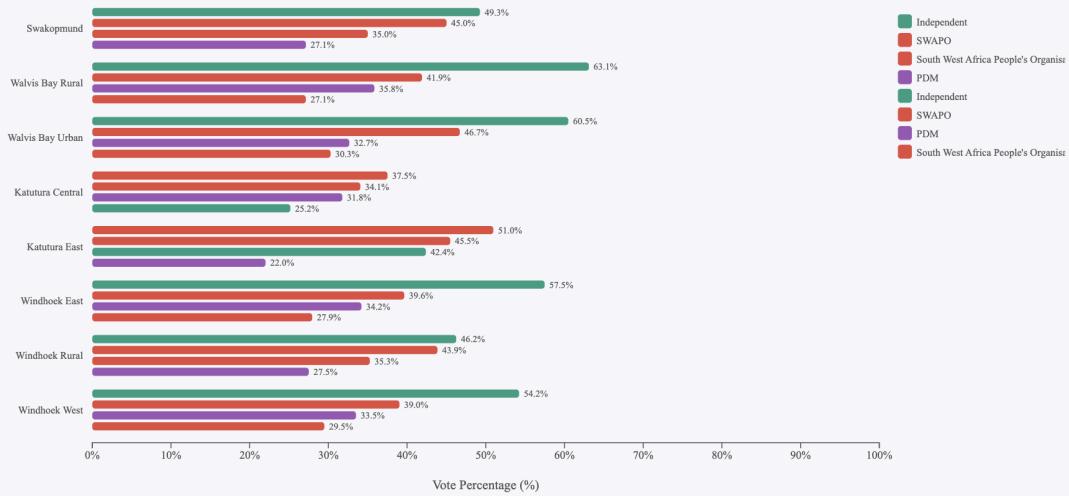
Shows the 15 most competitive constituencies (under 20% margin).

- Horizontal bars show vote shares for the top three parties and a “winnability” score

## Priority 2: Build Urban-Suburban Coalitions

Urban areas show extreme fragmentation with 6-7 viable opposition parties. Electoral pacts in these areas could consolidate the anti-SWAPO vote.

### Urban Fragmentation Analysis



#### Recommendation:

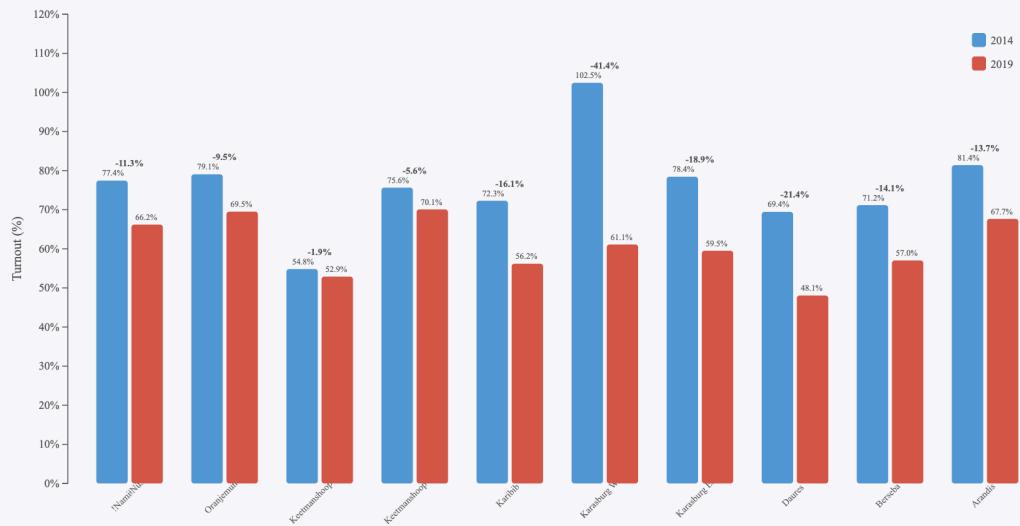
Form formal electoral pacts in urban constituencies where no single opposition party has clear dominance. Focus on Windhoek, Swakopmund, and Walvis Bay constituencies.

## Urban Fragmentation Chart

A bar chart comparing vote shares for major parties in urban constituencies like Windhoek, Walvis Bay, and Swakopmund.

### Priority 3: Address Turnout in Opposition Strongholds

#### Turnout Comparison: 2014 vs 2019



#### Potential Impact:

If 2019 turnout in opposition strongholds had matched 2014 levels, an estimated 12,000–15,000 additional opposition votes could have been captured, potentially flipping 2–3 close constituencies.

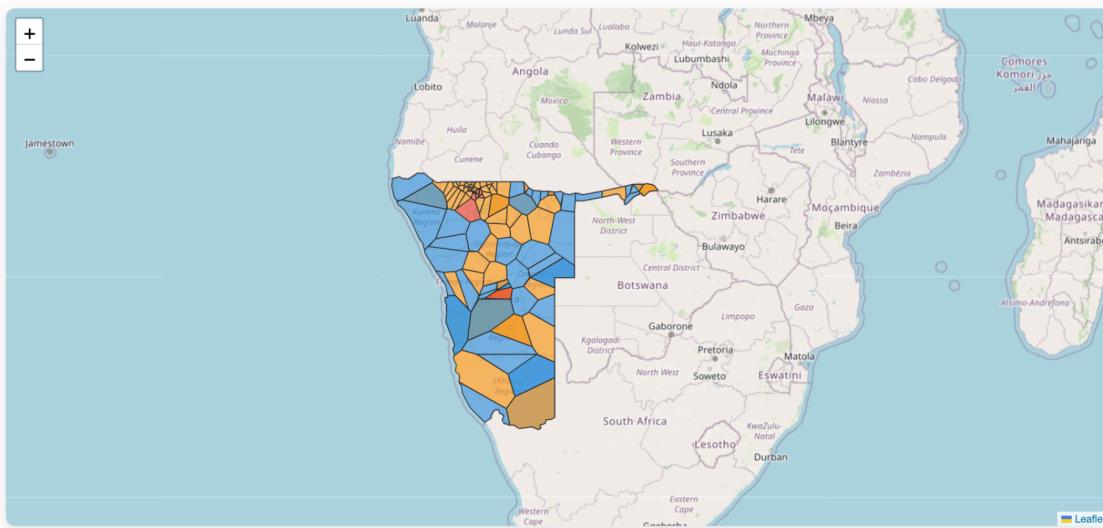
#### Turnout Change (2014 vs 2019)

Bar chart comparing voter turnout between 2014 and 2019 in opposition strongholds.

- Shows that lower turnout in 2019 cost the opposition an estimated 12,000–15,000 votes

## Projected Results: Coordinated Opposition Strategy

This map shows the potential electoral landscape with a fully coordinated opposition strategy.



### Conclusion:

Unity in Central Namibia is not just aspirational—it is the mathematically demonstrable path to building genuine opposition power. The data shows that fragmentation, not popular will, is the primary obstacle to breaking SWAPO's dominance.

## Projected Results Map

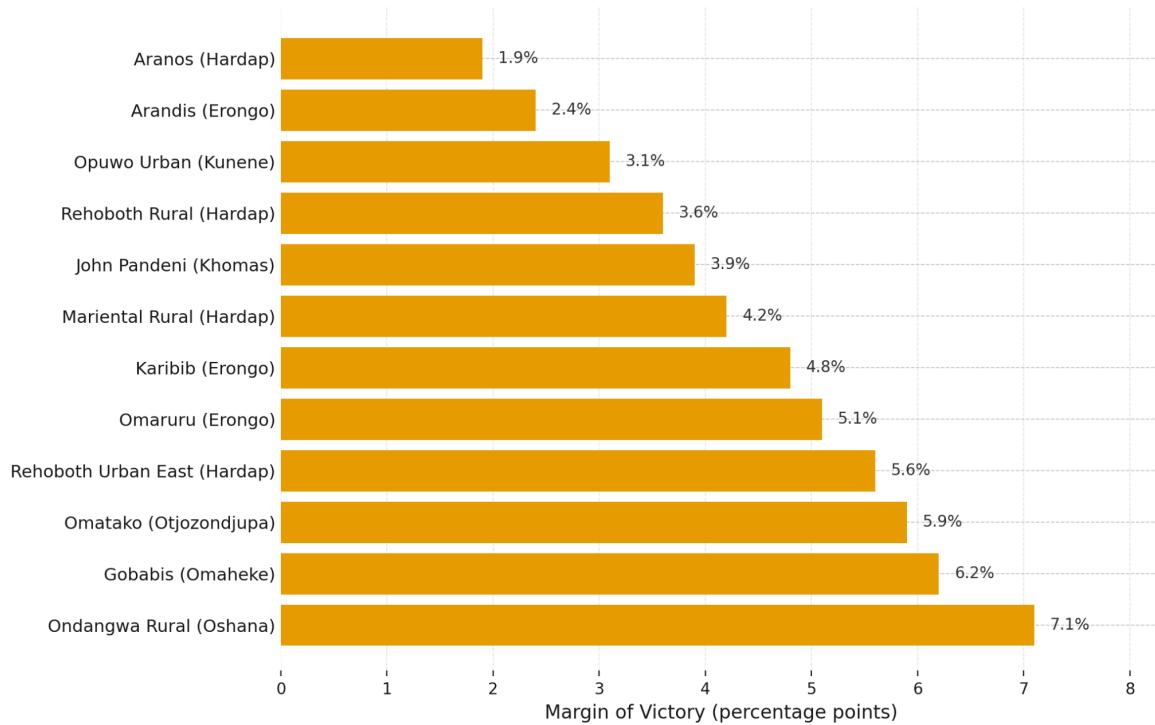
A final map showing what a fully coordinated opposition strategy could look like.

## Website Structure and Storytelling:

### Additional Visualization Drafts:

- Additional #1

### Closest Races (2019)

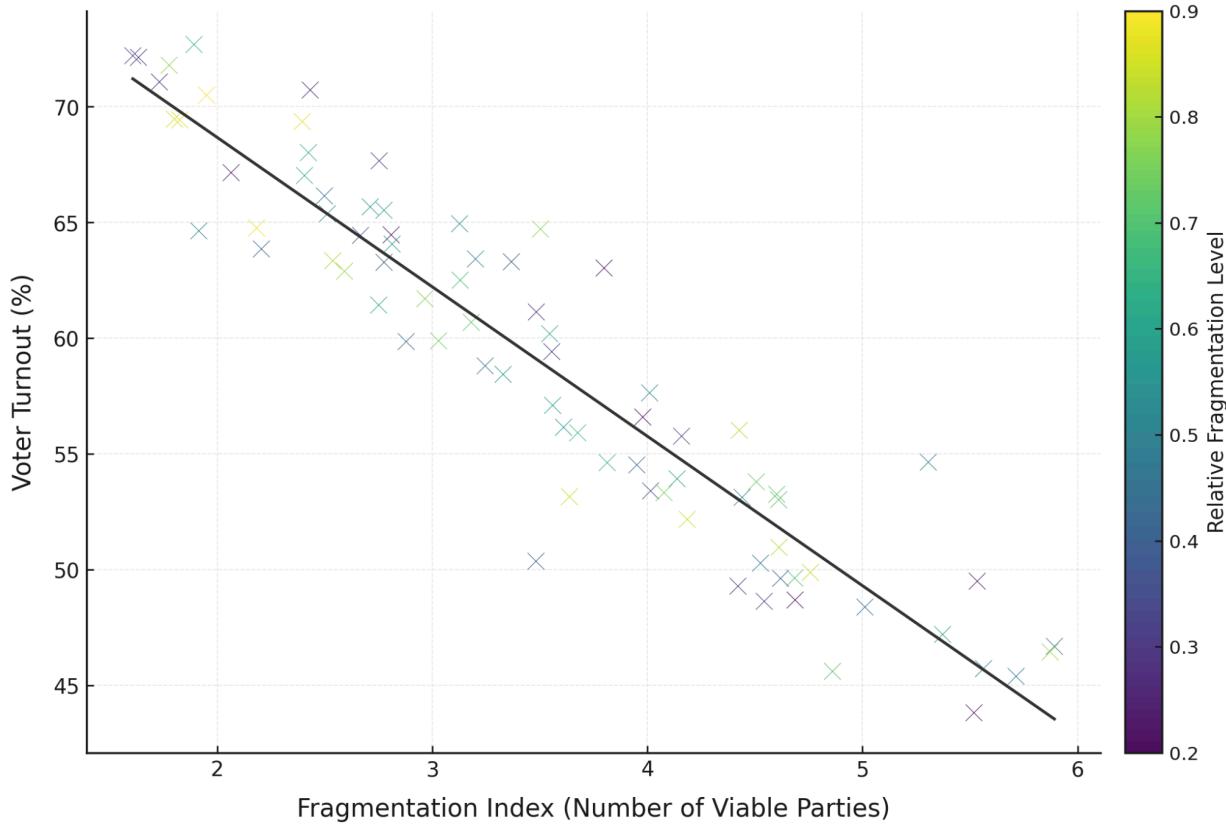


This Closest Races (2019) (used GPT to polish the design) chart shows the twelve constituencies with the smallest margins of victory in Namibia's 2019 election. Each bar represents one constituency, ranked from the tightest to the widest race. The layout makes it easy to see where elections were most competitive and where small changes in voter support could have changed the outcome. It gives a clear picture of which areas, especially in central and southern Namibia, are the most strategic for opposition parties in future elections

In the future, this chart could become interactive to make exploration easier, things like hovering over a bar could show extra details like total votes, turnout, and the top two parties. Clicking a bar could highlight that same area on the map or show how it changed between 2014 and 2019. Sorting or filtering options could also be added so users can view results by region or competitiveness. These features would make the data feel more alive and help viewers connect numbers to real places and outcomes.

- Additional #2

## Fragmentation vs Turnout (2019)



The Fragmentation vs. Turnout (2019) chart (used GPT to polish the design) shows the relationship between how divided a constituency's vote was and how many people turned out to vote. Each dot represents one constituency, with the x-axis showing how many viable parties competed and the y-axis showing voter turnout.

The message is evident from the downward slope: turnout decreases with increasing vote fragmentation. This implies that voters may feel their choices are less important and choose to stay at home when there are too many parties competing.

Thai can become interactive eventually in one main way: by clicking on a point, you can highlight that region on the main map or see how its attendance changed between 2014 and 2019. Users could also compare various elections or view patterns by region using a straightforward dropdown menu.

### Website Structure:

Our website is designed as a simple, scroll-based story that takes viewers through the main ideas step by step: “*Why does a party losing support keep winning?*” Each section focuses on one idea and includes either a map or chart, followed by a short explanation. Even though some of the visuals are still placeholders, the structure already shows how the final site could potentially look. It moves naturally from the problem, to the data, to the solution, using clean spacing and an easy-to-read layout that is consistent across the page.

### **Storytelling:**

The website and story starts with SWAPO’s ongoing success, then explains how regional differences, vote splitting, and low turnout all play a role. Each visualization adds another piece to the puzzle, helping the viewer see the bigger picture. By the end, the takeaway is clear: unity among opposition parties could shift political power in Namibia

### **Innovative View Design:**

Our most unique visualization is the Coordination Scenario Map, which shows how election results would change if opposition parties joined forces. Viewers can switch between the “fragmented” and “coordinated” maps to see how power could shift across Namibia. It helps people understand not just what has happened, but what could happen if parties worked together.

## Milestone 8:

- **Individual exploration:**
  - Catherine:
    - Geographic polarization of electoral decline
      - SWAPO experienced dramatic vote share losses in southern and central regions between 2014 and 2019 while maintaining strongholds in northern regions.
    - Turnout paradox
      - Voter turnout declined nationally from 2014 to 2019, but opposition parties gained vote share. In regions where SWAPO lost the most support (like Khomas) turnout dropped. This suggests that SWAPO’s losses might have been due to low turnout instead of party switches.
    - Urban-Rural divide in party competition
      - Urban constituencies (Windhoek, Walvis Bay) show higher party fragmentation with 4-6 viable parties while rural northern constituencies remain one-party dominated. This suggests that opposition parties have limited pathways to power without breaking into rural areas.
  - Matteo:
    - Competitive constituencies are clustered

- The constituencies with the smallest margins of victory (under 20%) are geographically concentrated in central Namibia and the south, while the north shows landslide victories. It is very likely that SWAPO's historical dominance decreased most in areas with more diverse economies and more organized opposition. This means opposition parties should focus resources on a defined geographic area rather than spreading efforts nationwide.
- Constituency size doesn't predict turnout
  - There's no clear relationship between registered voters and turnout rates, but there are notable outliers, some small constituencies show unusually high turnout (>100%, likely data issues) while some large urban constituencies show surprisingly low turnout. These anomalies warrant investigation since there should be some semblance of a pattern amongst them.
- Fragmentation correlates with SWAPO weakness
  - Regions with the highest party fragmentation scores (Khomas, Erongo) are exactly where SWAPO lost the most ground. However, this fragmentation may actually benefit SWAPO if opposition parties split the anti-incumbent vote rather than coordinating
- Nathaniel:
  - Regional party dynamics show emerging opposition patterns
    - Between 2014 and 2019, the opposition vote consolidation varies dramatically by region. In some regions (Hardap, ||Karas), the opposition coalesced behind PDM, while in others (Khomas), the vote fragmented across multiple parties. This suggests different opposition strategies are working in different contexts.
  - The 2019 presidential race signals future party realignment
    - Independent candidate Panduleni Itula captured 248,552 votes (nearly 30% of votes cast) by running *against* SWAPO despite being a former party member. His support was concentrated in specific regions, suggesting a potential blueprint for opposition parties to peel away SWAPO supporters rather than just mobilizing new voters.
  - Second-place parties matter more than vote share
    - Looking at who came in second place reveals more about electoral dynamics than just SWAPO's vote share. In 2019, different opposition parties won second place in different regions, showing no unified opposition front—which may be the key factor preventing SWAPO's defeat despite their declining support.

**Main insight following discussion:** Opposition parties can win constituencies and potentially national power by consolidating their fragmented support in central and southern Namibia, where SWAPO has already lost its majority but opposition votes remain divided across multiple parties.

**Explanation of main insight:**

This insight is actionable as it directly suggests a political strategy for the opposition parties, it is supported by visualizations that show SWAPO losing ground; opposition fragmentation; and geographic clustering of competitive races, and finally it is both surprising and time-sensitive with recent election data soon to inform future work.

## **Data Storyboard:**

- We attached a zip file containing a storyboard to open in LiveServer on Canvas!

**HOOK:** "Why Does a Party Losing Support Keep Winning?"

Visualization used: Side-by-side maps showing SWAPO vote share 2014 vs 2019

- Show dramatic color shift (red to lighter pink) across southern regions
- Text overlay: "SWAPO lost 7-12% vote share in 6 of 14 regions... but still won 12 of 14"
- Purpose: Create intrigue—how can a party lose so much support but maintain power?

## **Rising insights:**

Insight 1: Geographic Polarization

*Visualization: Interactive choropleth map showing SWAPO vote share change by region*

- Click through regions to see detailed constituency results
- Highlight the "Red Line" boundary showing historical north-south divide
- Purpose: Establish that SWAPO's losses are geographically concentrated, not uniform

Insight 2: The Opposition Fragmentation Problem

*Visualization: Stacked bar charts showing vote distribution in "flipped" constituencies*

- Show 6 constituencies where SWAPO lost majority
- Bars divided by party (PDM, LPM, IPC, ITULA, etc.)
- Purpose: Reveal that in regions where SWAPO lost, no single opposition party capitalized—the vote split many ways

Insight 3: Where Races Are Competitive

*Visualization: Margin of victory heatmap*

- Show which constituencies had <15% margin (true swing districts)
- Color code by region to show clustering

- Purpose: Demonstrate that winnable seats concentrate in central/southern Namibia, not scattered nationwide

Insight 4: Successful Opposition Consolidation Examples

*Visualization: Before/after comparison in Hardap region*

- 2014: Opposition vote split 5 ways
- 2019: PDM consolidated to 52% in some constituencies
- Purpose: Prove the strategy works—show concrete example where coordination led to opposition victory

If possible, we will also involve recent 2024 election data if it is released within the timeline of this final project!

**Main Message:** Unity in Central Namibia Is the Path to Power

Visualization: Scenario modeling graphic

- Map showing current results (fragmented opposition)
- Map showing "what if" scenario where top 2 opposition parties coordinated
- Calculate: "With coordination, opposition would win 4 additional constituencies + gain 8 more competitive seats"
- **Purpose:** Make the strategic implication crystal clear with specific numbers

**Solution:** Three Strategic Priorities for Future Elections

*Visualization: Action roadmap with supporting data*

**Priority 1: Focus geographic targeting**

- List of 15 most winnable constituencies (all in central/south)
- Small multiples showing each constituency's party breakdown
- **Recommendation:** Concentrate resources on these constituencies rather than attempting nationwide campaigns.

**Priority 2: Build urban-suburban coalitions**

- Show urban fragmentation data (6-7 parties viable in Windhoek)
- Highlight suburban constituencies where 2-party coordination would flip seats
- **Recommendation:** Formal electoral pacts in urban areas where no single opposition party dominates

### **Priority 3: Address turnout in opposition strongholds**

- Show constituencies where opposition won but turnout dropped
- Calculate potential additional votes if turnout matched 2014 levels
- **Recommendation:** Mobilization in existing opposition areas is as important as conversion in swing districts

*Final visual: Updated map showing projected results with coordinated opposition strategy*

## Milestone 7:

### **Paragraph Reflection:**

We chose the sketch IDs 1, 2, 3, and 4. Sketches 1 and 2 have the highest number of votes. The reason why we chose these sketches is because they answered the greatest number of questions. We want to address as many questions as possible in our initial data visualizations for this project, so these four sketches were the logical approach. The most difficult to visualize using digital visualization for our final project website will be Sketch 1, as we will have to learn to use the geojson package. We have sourced the dataset in geojson format, so this will now be easier for us to work with! We are excited to see where these sketches take us and how we can make them interactive on a website. One idea we have is to make the map for Sketch ID 1 appear in the shaded color regions for vote share, but also have an interactive option where one can also look at the country's topography. We believe that this is doable using geojson. These sketches are exciting to analyze and we are excited to continue with this process!

Sketch ID	Question ID	Author	<b>Group Vote</b>
1	1, 4, 5	CM	4
2	1, 3, 4, 5	CM	4
3	2, 5	CM	3
4	1, 3, 6	MW	3

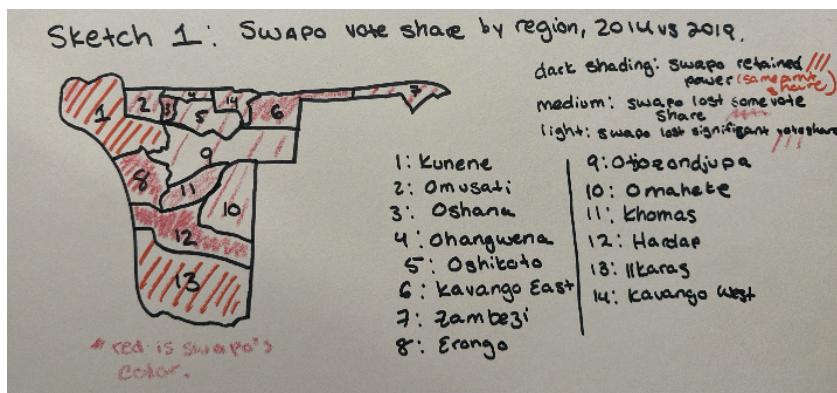
5	7	MW	1
6	3	CM	1

### Question IDs:

1. How did SWAPO vote share by region change between 2014 and 2019?
2. Knowing that SWAPO lost votes between 2014 and 2019, how was opposition vote share divided in majority-opposition-party regions in the 2019 election?
3. How does voter turnout correlate with SWAPO performance?
4. How did SWAPO support change between 2014 and 2019 across all 14 observed regions?
5. How have the North and South areas of Namibia behaved differently in the 2014 and 2019 elections, and is there a geographic pattern?
6. How do closer elections correlate with voter turnout?
7. Which regions exhibit the greatest party fragmentation?

### Sketch IDs:

#### Sketch ID 1, CM

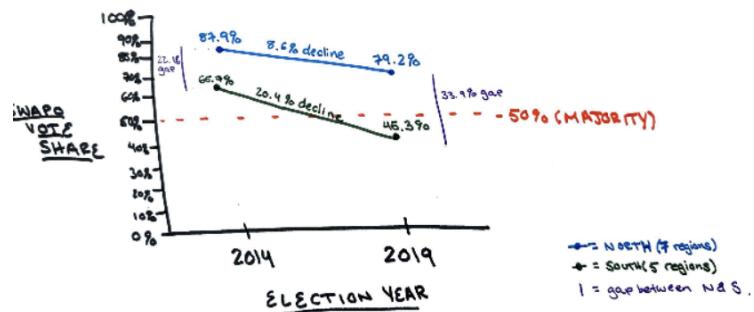


#### Sketch ID 2, CM

Sketch 5: Geographical voting behavior (North vs. South).

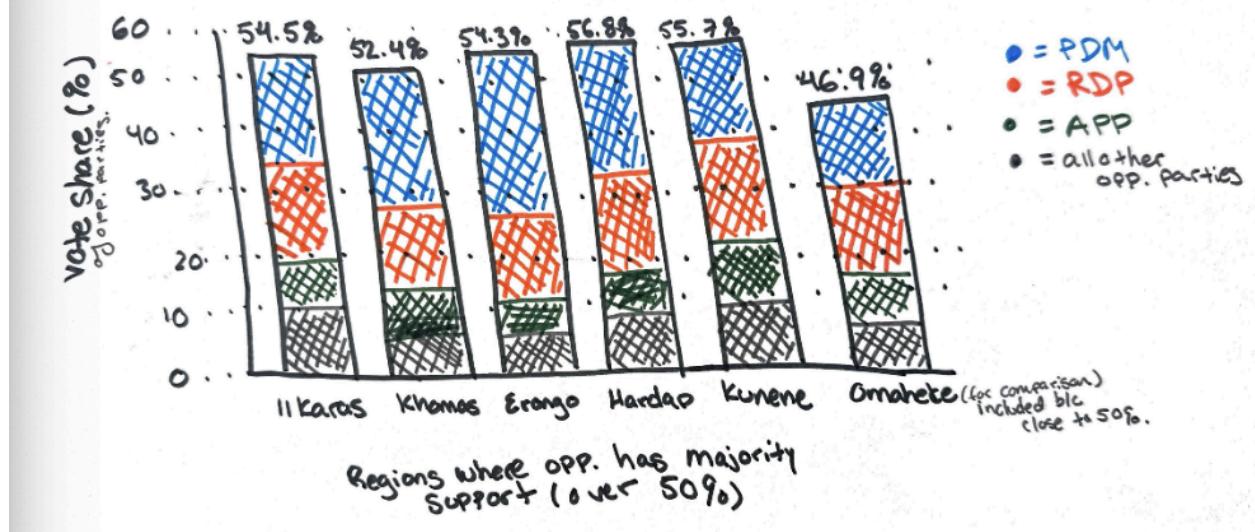
Answering: How have the North & South areas of the country behaved, & is there a geographic pattern?

Answer: Yes! We should look into the "Red Line", which divides the country.

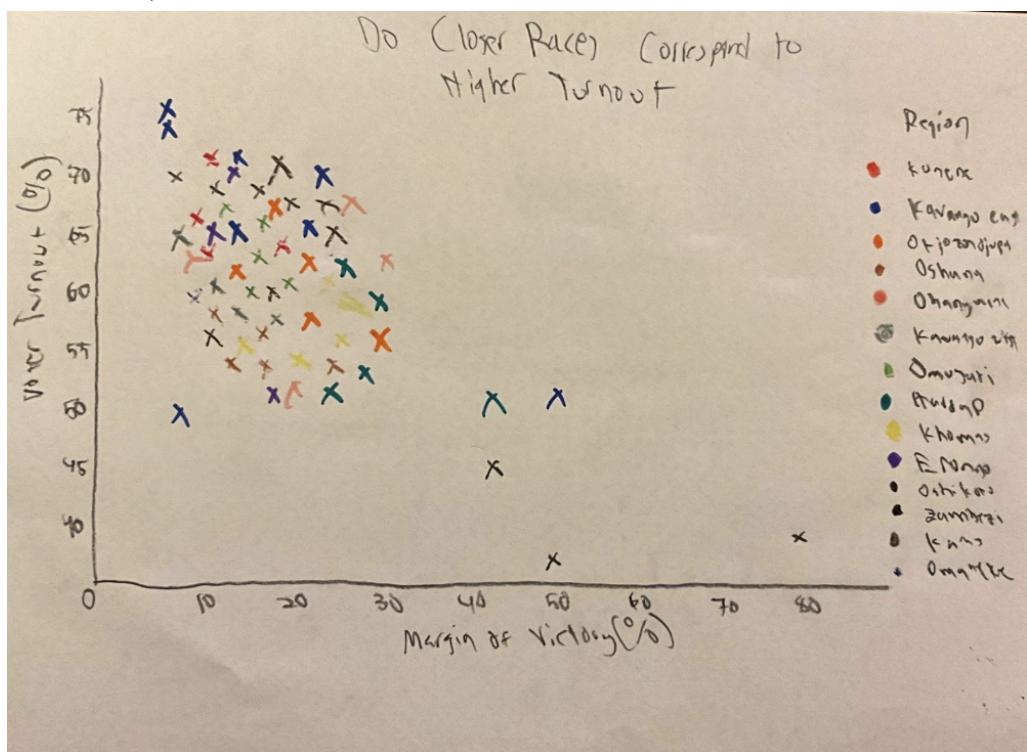


Sketch ID 3, CM

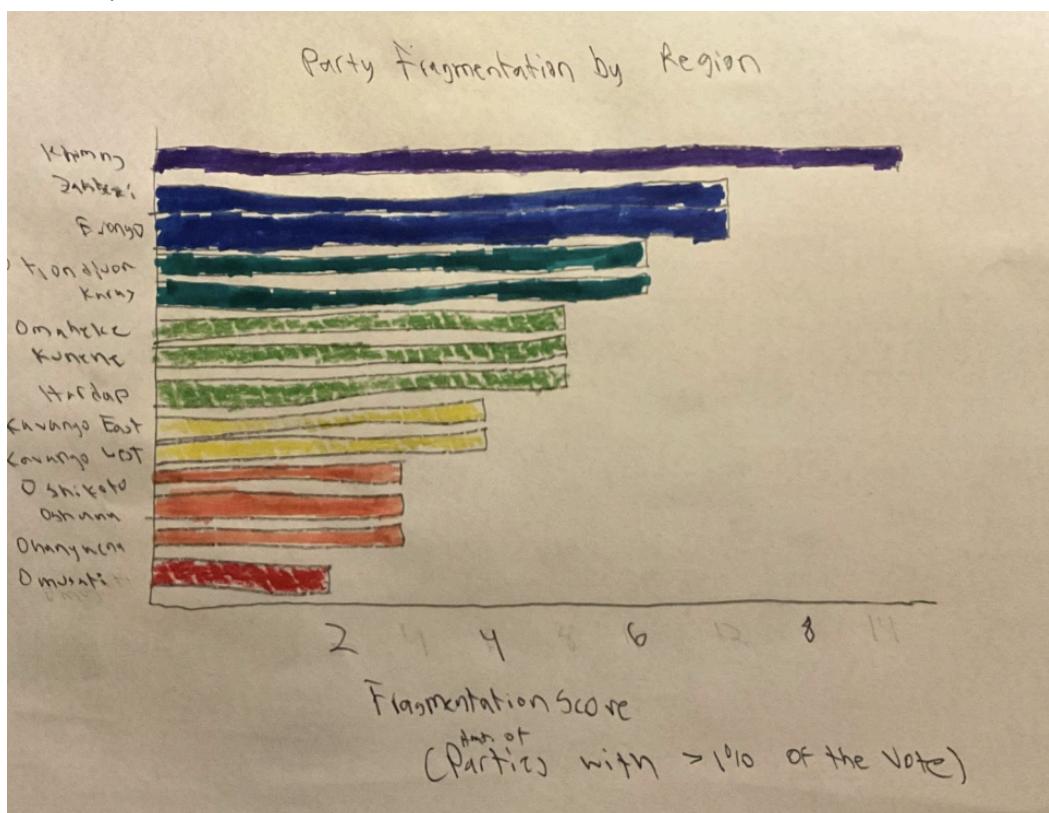
Sketch 2: Opposition Vote Share by Region, 2019 Election.



Sketch ID 4, MW



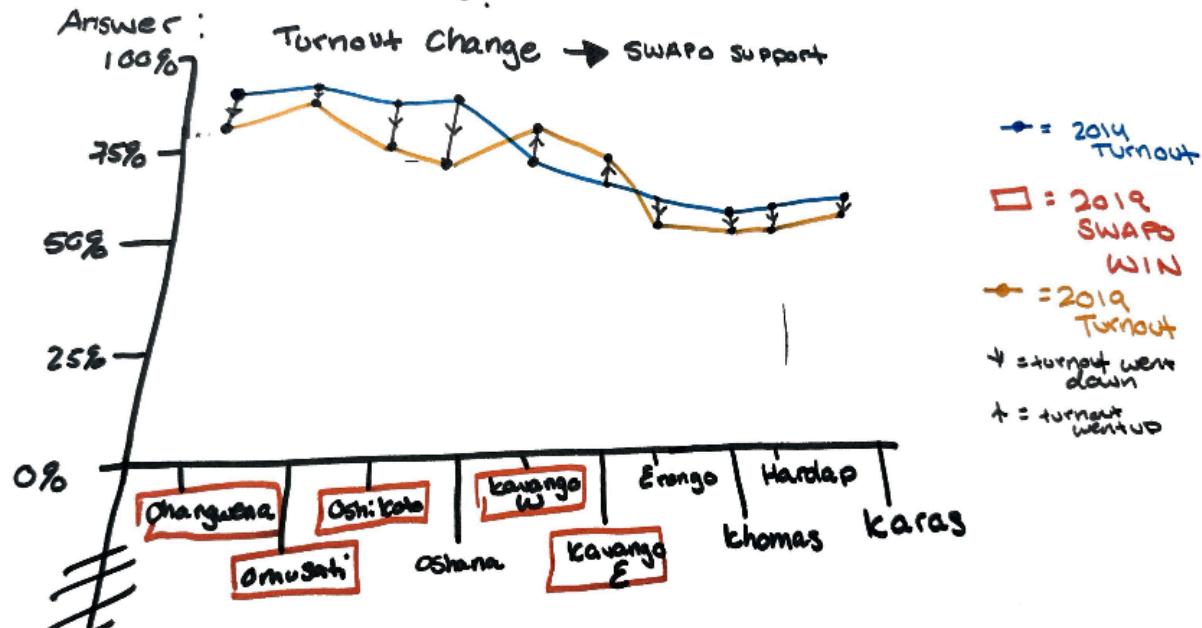
Sketch 5, MW



## Sketch 6, CM

Sketch 3: Turnout vs Party Performance

Answering: Does voter turnout correlate w/ SWAPO performance?



## Milestone 6:

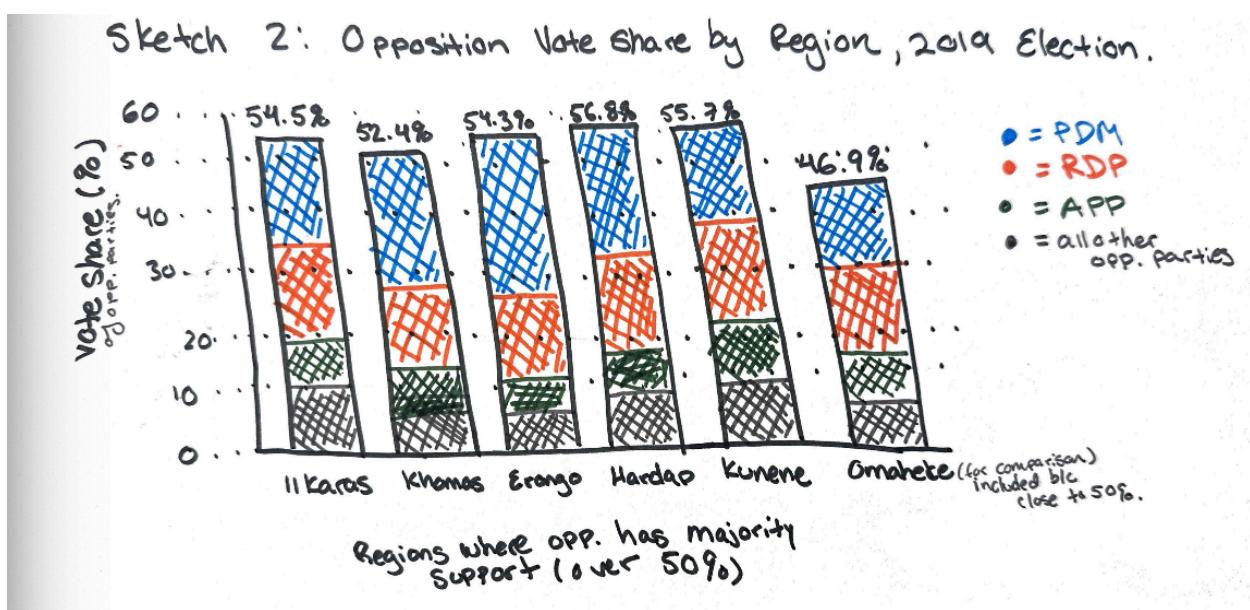
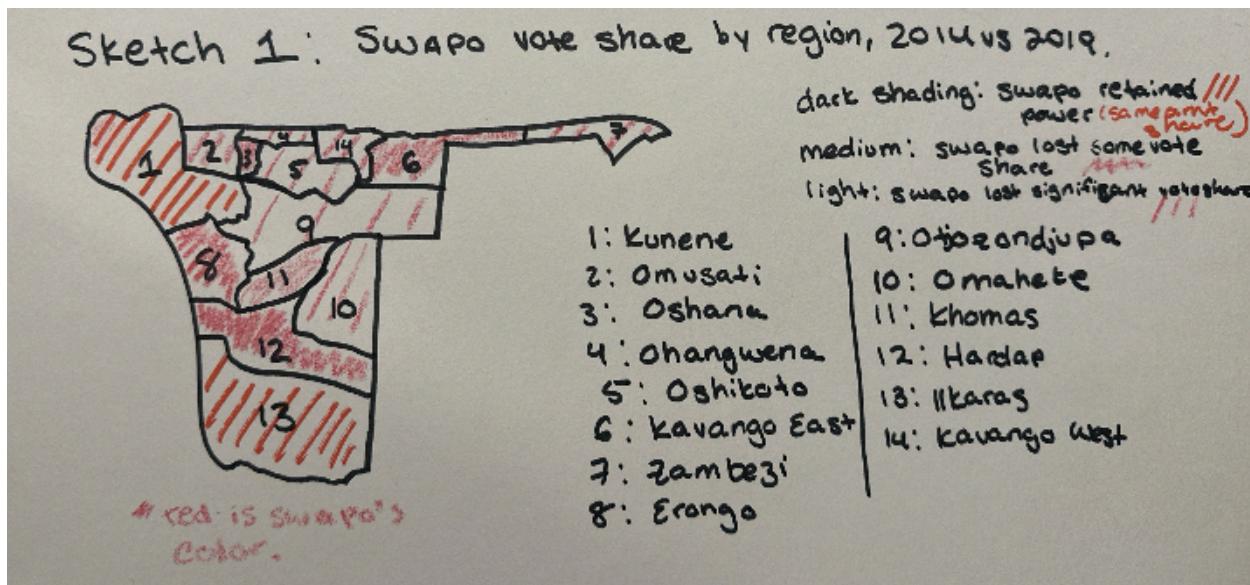
Catherine's Sketches:

Please find the [PDF linked here](#) for easy viewing.

Questions I looked to answer with these sketches:

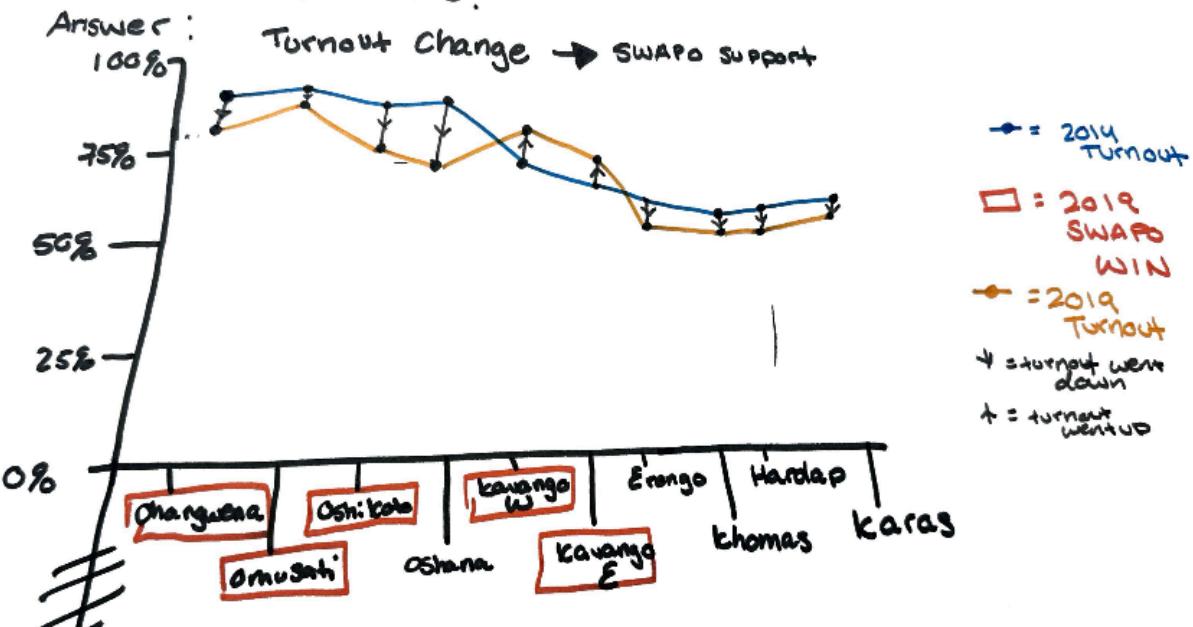
- Sketch 1: How did SWAPO vote share by region change between 2014 and 2019?
- Sketch 2: Knowing that SWAPO lost votes between 2014 and 2019, how was opposition vote share divided in majority-opposition-party regions in the 2019 election?
- Sketch 3: How does voter turnout correlate with SWAPO performance?
- Sketch 4: How did SWAPO support change between 2014 and 2019 across all 14 observed regions?
- Sketch 5: How have the North and South areas of Namibia behaved differently in the 2014 and 2019 elections, and is there a geographic pattern?

I also share the images below (but the PDF is slightly clearer due to file size):

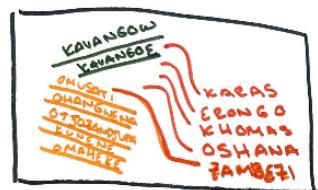


### Sketch 3: Turnout vs Party performance

Answering: Does voter turnout correlate w/ SWAPO performance?



### Sketch 4: Dear - Data Inspired, SWAPO support across 14 regions



RED: steepest drop between 2014 & 2019  
ORANGE: gradual decline between 2014 & 2019  
GREEN: gradual incline between 2014 & 2019

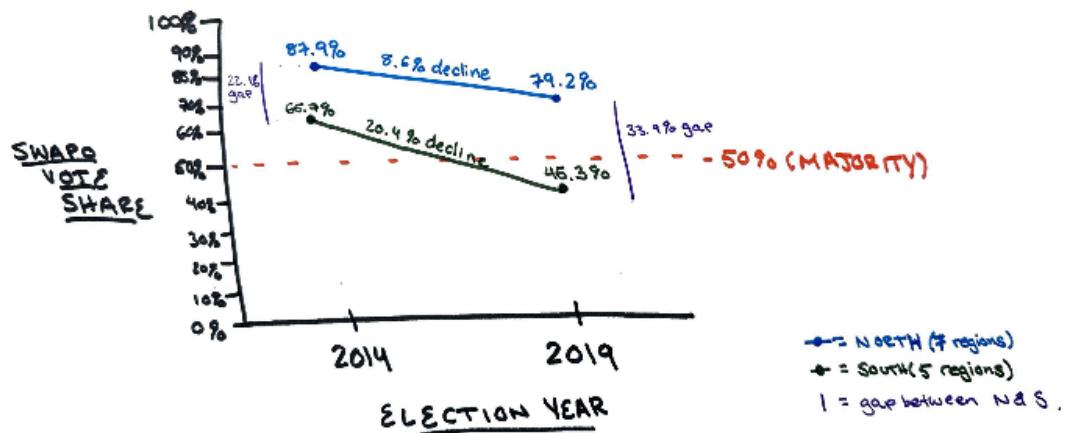


I envision a dear-data inspired page on the FP website to use interesting visualizations in addition to our more 'serious' ones! These would be interactive dear-data visualizations online.

Sketch 5: Geographical voting behavior (North vs. South).

Answering: How have the North & South areas of the country behaved, & is there a geographic pattern?

Answer: Yes! we should look into the "Red Line", which divides the country.

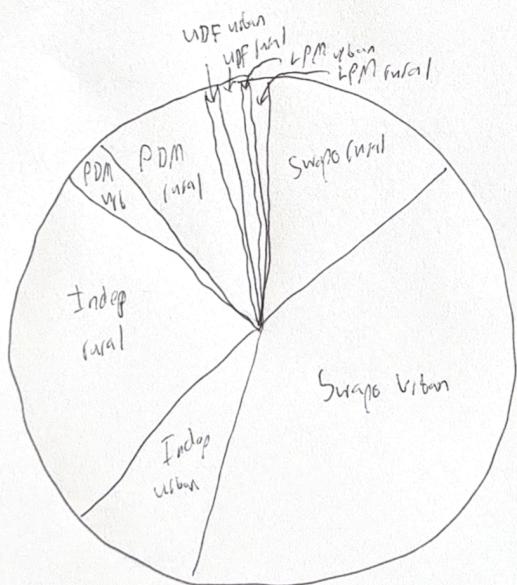


Nathaniel Sketches ([high res link](#)):

How do levels of fragmentation and competitiveness differ between urban and rural regions?

Percentage breakdown if hover over pie chart section

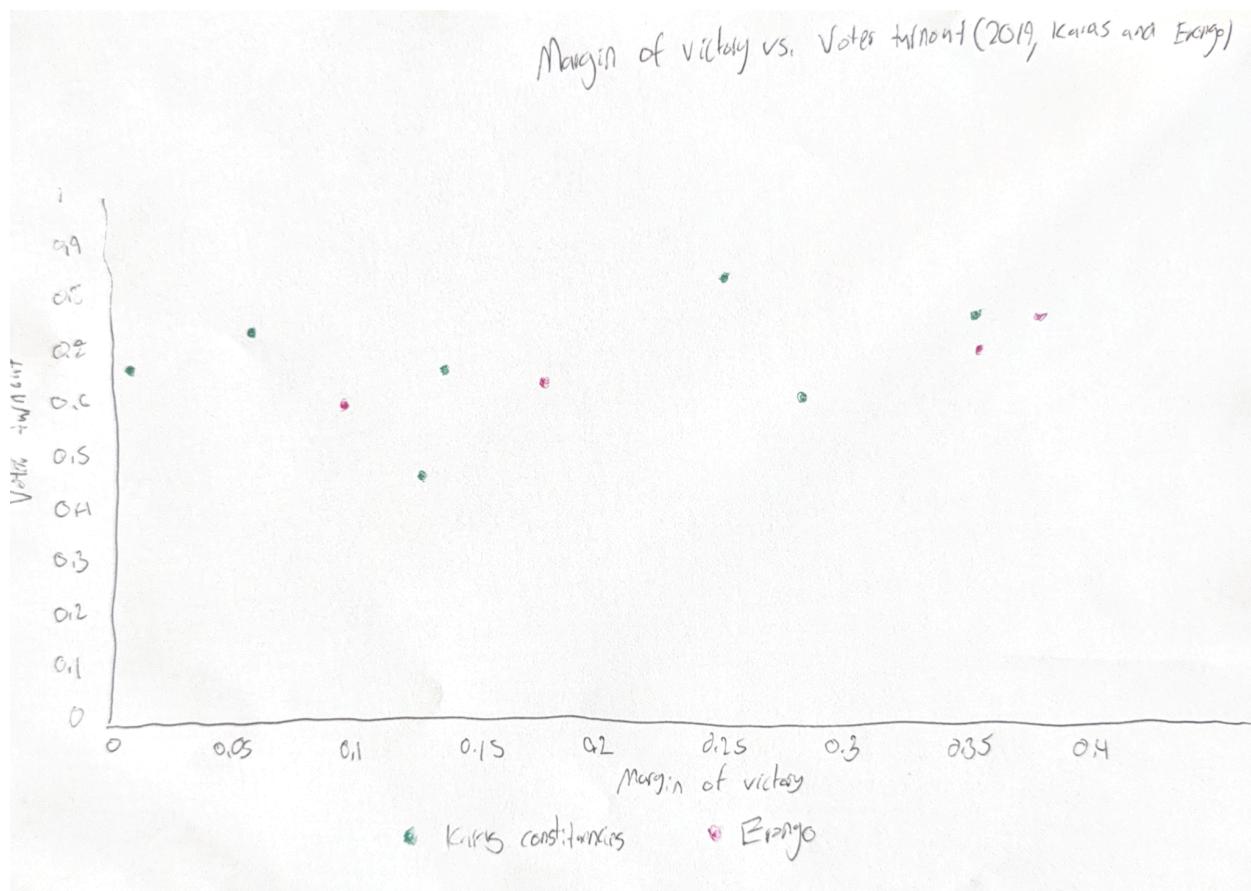
Vote share in urban vs Rural (ICARUS, 2010)



Do closer races tend to be associated with higher voter turnout?

Further click-throughs if hover over points

In



Which regions exhibit the greatest party fragmentation, and how might that fragmentation create opportunities for opposition parties?

Pie chart will have click throughs with percentages



Which constituencies were won by the smallest margins, and do these competitive areas cluster in certain regions?



Which regions exhibit the greatest party fragmentation, and how might that fragmentation create opportunities for opposition parties?



## Matteo Wakeman Sketches:

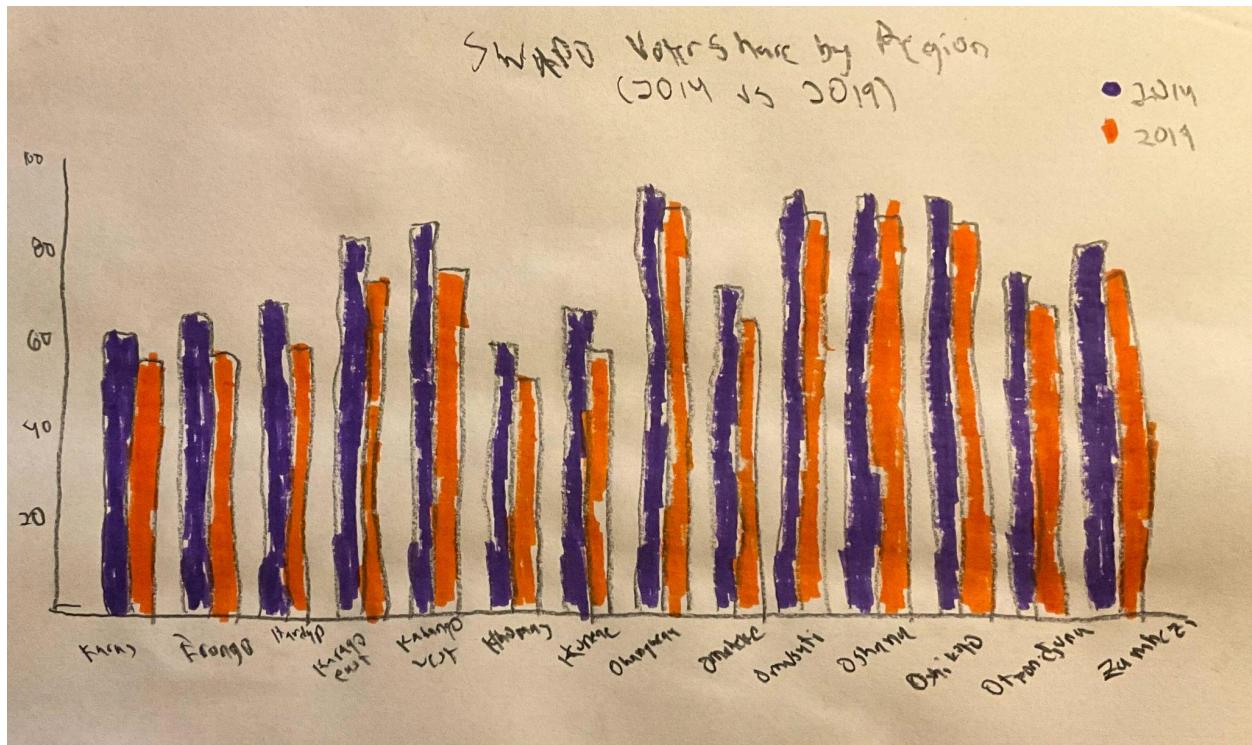
Sketch 1: SWAPO Voter Share by Region from 2014 to 2019

### Question answered:

How has SWAPO (the ruling party) lost vote share over time across regions?

### Insight shown:

By comparing side-by-side bars for 2014 (purple) and 2019 (orange), this sketch highlights SWAPO's declining vote share in nearly every region, especially Khomas, Erongo, and Hardap. It makes clear where support remained strong (Omusati, Oshana) versus where it weakened most.



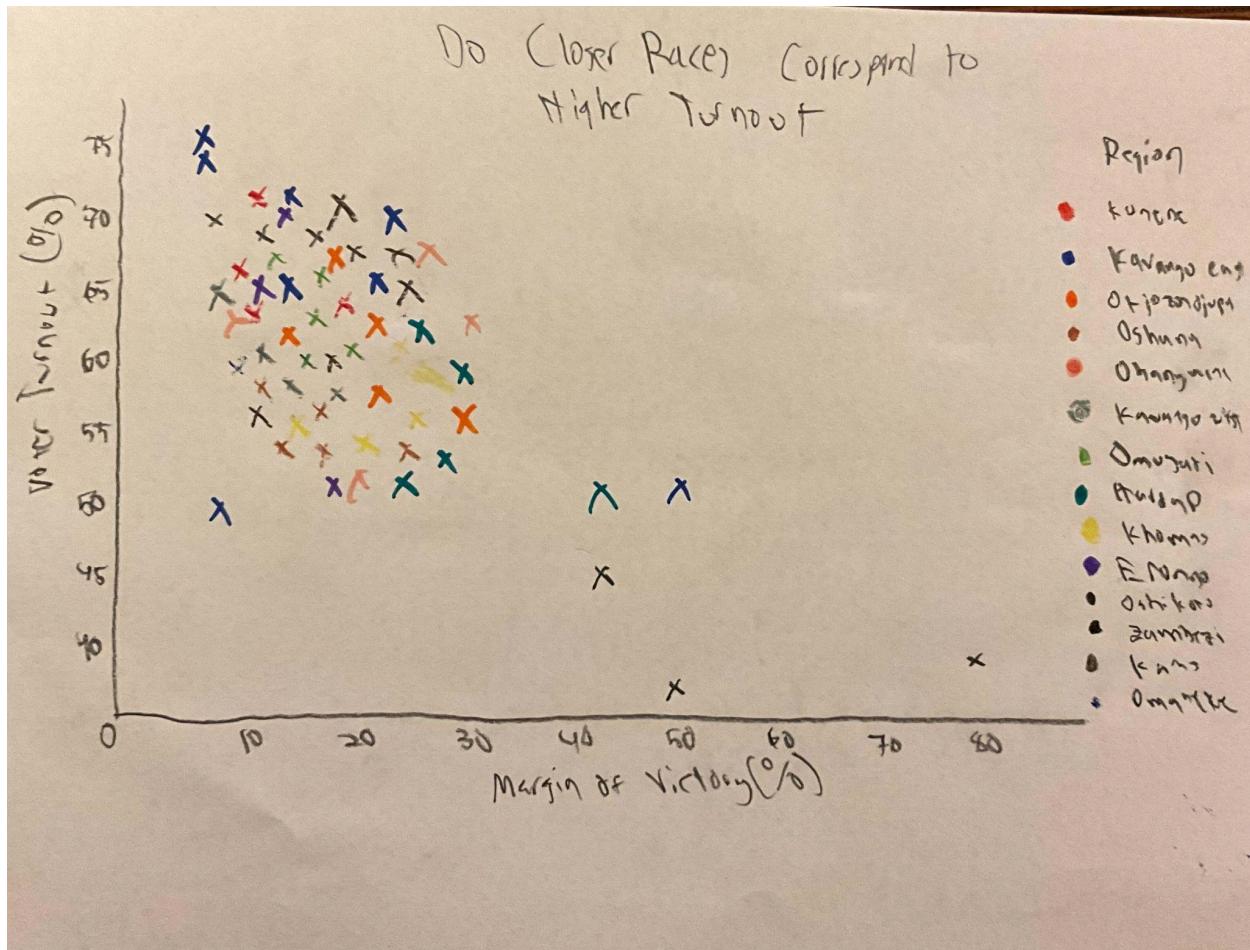
Sketch 2: Do Closer Races Correspond to Higher Voter Turnout

**Question answered:**

Do constituencies with smaller margins of victory show higher voter turnout?

**Insight shown:**

Each point represents a constituency. A loose downward pattern suggests that as the margin of victory widens, turnout falls—meaning competitive races tend to motivate more voters to participate. The color coding by region adds another spatial dimension.



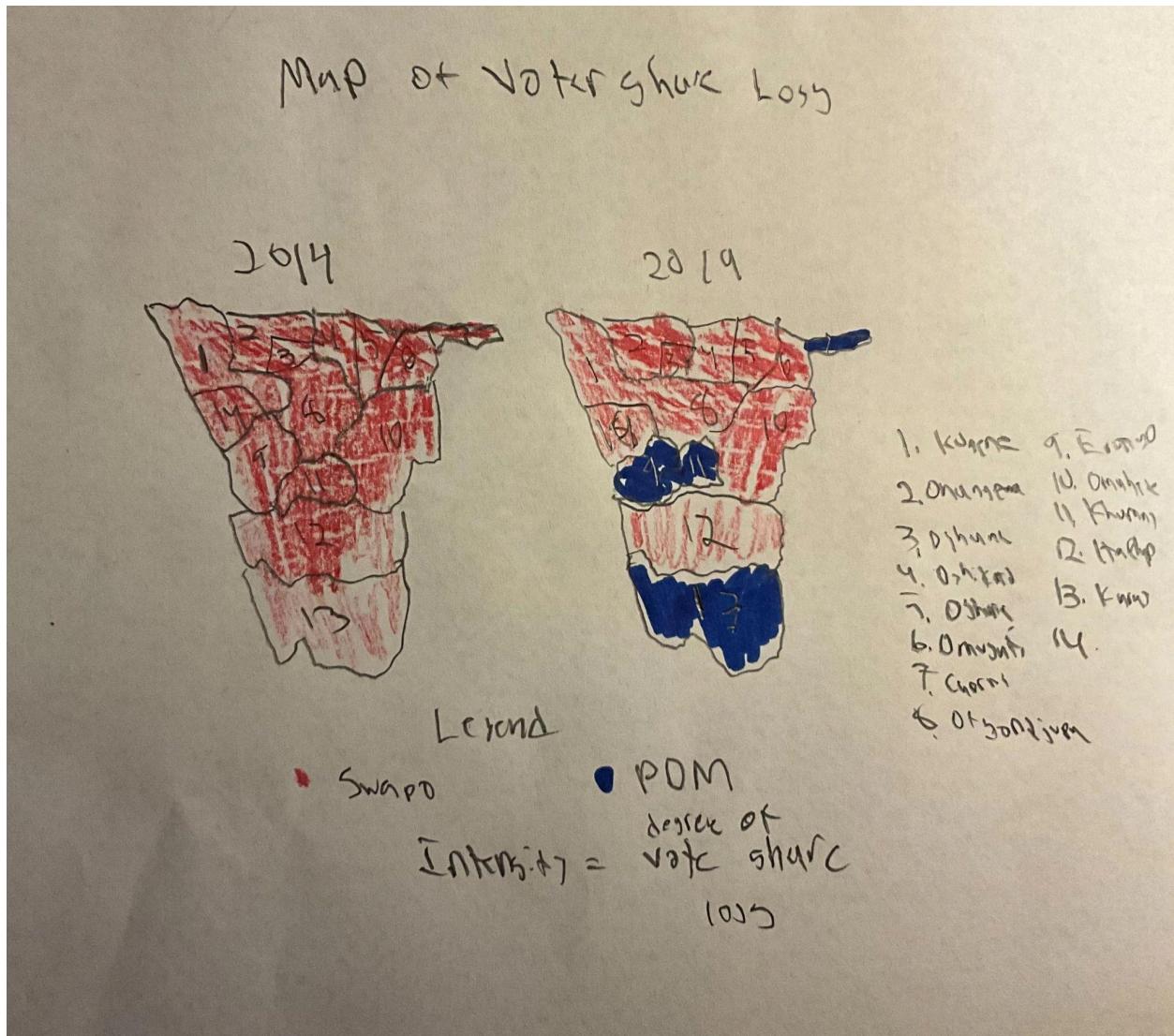
Sketch 3: Map of Voter Share Loss

**Question answered:**

Where in Namibia did SWAPO lose the most support between 2014 and 2019, and do these losses cluster geographically?

**Insight shown:**

The two side-by-side maps use shading (red = SWAPO, blue = PDM) to show how formerly solid SWAPO regions in the south and center turned more competitive in 2019. This visual clearly conveys spatial clustering of vote-share loss, particularly around Windhoek and southern constituencies.



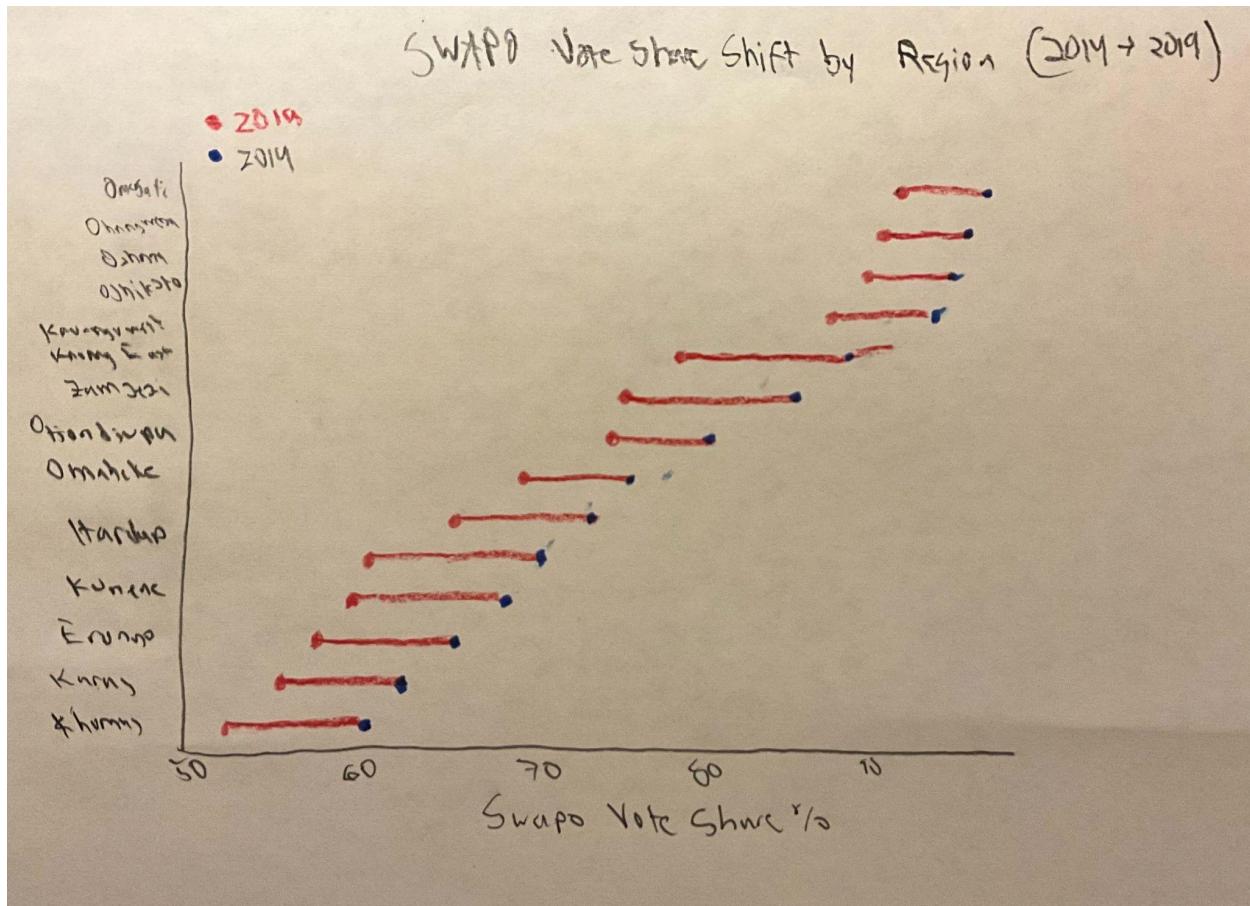
Sketch 4: Voter Share Shift Per Region

**Question answered:**

By how much did SWAPO's vote share shift between 2014 and 2019 in each region?

**Insight shown:**

This slope chart measures change rather than absolute levels. Each line connects 2014 (blue dot) to 2019 (red dot), showing direction and magnitude of movement. Steeper downward lines indicate regions where SWAPO lost the most support — a clean way to quantify decline while preserving comparative scale across all regions.



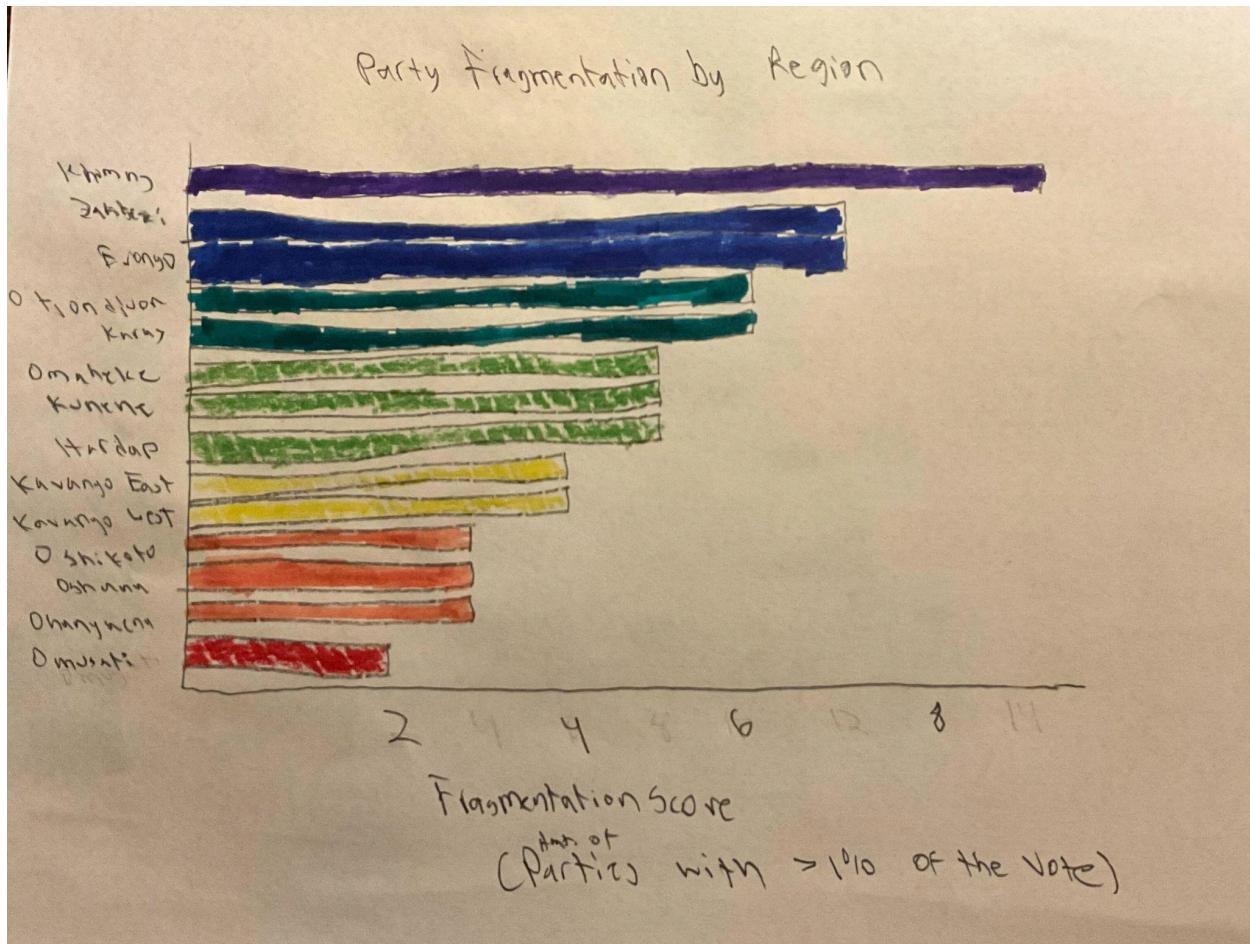
Sketch 5: Party Fragmentation By Region

**Question answered:**

Which regions exhibit the greatest party fragmentation (most diverse political competition)?

**Insight shown:**

Horizontal bars measure the number of parties earning > 1 % of votes. Darker blue regions such as Khomas and Erongo stand out as highly fragmented, while light-colored northern regions remain dominated by SWAPO. This shows how urbanization correlates with multiparty competition.



## Milestone 5:

### Part 1, Finalize and document your data sources:

Dataset: Namibian election dataset, 2014 and 2019. We have confirmed that we will use this dataset which was generated by the Namibian government and made public.

File format: csv

No usage restrictions.

## **Part 2, Create a data inventory:**

- a. Country:
  - i. Geographical, categorical. All values are collected from the same country,: Namibia.
- b. NAME E1
  - i. Region of vote, geographical.
- c. NAME E2
  - i. Subregion of vote, geographical.
- d. Year
  - i. Temporal.
- e. Election Type
  - i. Categorical.
- f. Party
  - i. The winning party in the election. Categorical.
- g. Party B
  - i. The runner-up party in the election. Categorical.
- h. Votes
  - i. Quantitative, collected by the Namibian government.
- i. Total Votes
  - i. Quantitative, collected by the Namibian government.
- j. Registered Voters
  - i. Quantitative, collected by the Namibian government.
- k. Turnout\_Reg (Voter Turnout)
  - i. Quantitative, calculated by the Namibian government.
- l. Latitude
  - i. Geographic, collected via Google.
- m. Longitude
  - i. Geographic, collected via Google.

## **Part 3, Data Exploration:**

Initial prompt to Claude: generate a basic statistical summaries and detect outliers in addition to missing values for all non-geographic variables.

Since it is a large dataset, it took some time for Claude to generate a response. Most notably, Claude detected what we had previously mentioned: that there were more votes than registered voters in some regions.

**Claude's Report (after back-and-forth discussion of desired summary statistics):**

## Dataset Overview

- **Total Records:** 5,072
- **Columns:** 17
- **Years Covered:** 2014, 2019
- **Election Types:** Legislative (lower house), Presidential
- **Geographic Coverage:** 14 regions, 125 constituencies
- **Political Parties:** 30 unique parties

### 1. Data Structure

#### Categorical Variables

- **country:** Namibia only (1 unique value)
- **NAME1:** 14 regions (Karas, Erongo, Hardap, Kavango East, Kavango West, Khomas, Kunene, Ohangwena, Omaheke, Omusati, Oshana, Oshikoto, Otjozondjupa, Zambezi)
- **NAME2:** 125 constituencies
- **year:** 2014, 2019
- **election\_type:** Legislative (lower house), Presidential
- **party:** 30 unique parties
- **party\_b:** Party abbreviations

#### Numeric Variables

Variable	Count	Min	Max	Mean	Median	Q1	Q3
votes	5,072	0	17,965	499.30	37	13	132
total_votes	5,072	1,442	21,587	7,004.93	5,941	3,614	9,564
reg	5,072	1,747	45,410	10,864.66	8,544	5,885	13,909
turnout_reg	5,072	0.29	1.38	0.67	0.68	0.55	0.78
latitude	5,072	-28.55	-17.00	-19.92	-18.15	-22.40	-17.77
longitude	5,072	13.24	24.63	17.33	16.91	15.71	18.49

### 2. Missing Values Analysis

- ✓ **No missing values detected** in any column. All 5,072 records are complete across all 17 variables.

### 3. Duplicate Analysis

- **Exact duplicate rows:** 0
- **Logical duplicates** (same year-constituency-party combination): 0
- Each record represents a unique party-constituency-year combination

### 4. Data Consistency Issues

#### Critical Issues Found

##### Issue 1: Turnout > 100% (180 records)

- **Problem:** 180 records show total\_votes exceeding registered voters (reg)
- **Impact:** Turnout rates range from 101.69% to 138.47%
- **Example:** Karasburg West (2014) - total\_votes: 5,044, reg: 4,923, turnout: 102%
- **Possible causes:**
  - Same-day registration not captured in reg field
  - Data entry errors
  - Mismatched registration data vs actual voting data

##### Issue 2: Zero Votes (48 records, 0.95%)

- **Problem:** 48 records have zero votes for a party
- **Consideration:** May be legitimate (party participated but received no votes) or a data quality issue.

#### No Issues Found

- All votes  $\leq$  total\_votes (valid)
- No negative values in any numeric field
- Turnout\_reg calculations are consistent with total\_votes/reg

### 5. Distribution Analysis

#### Records by Year and Election Type

- **2014:** 1,926 records (all Legislative)
  - Total votes cast: 888,606
- **2019:** 3,146 records

- Legislative: 1,815 records
- Presidential: 1,331 records
- Total votes cast: 1,643,842

#### Regional Distribution (Top 5)

1. Ohangwena: 504 records
2. Omusati: 504 records
3. Oshana: 462 records
4. Oshikoto: 462 records
5. Khomas: 420 records

#### Top 5 Parties by Total Votes

1. **South West Africa People's Organisation:** 1,248,864 votes
2. **GEINGOB, Hage** (Presidential): 464,129 votes
3. **ITULA, Panduleni** (Presidential): 248,552 votes
4. **Popular Democratic Movement:** 136,293 votes
5. **VENAANI, McHenry** (Presidential): 43,880 votes

*Note: Presidential candidates appear as "parties" in 2019 data*

## 6. Outlier Detection

### Votes Outliers

- **Upper bound (IQR method):** 310.5 votes
- **Outliers identified:** 837 records (16.5%)
- **Highest vote counts:**
  1. 17,965 - SWAPO, Outapi (2014)
  2. 15,330 - SWAPO, Outapi (2019)
  3. 14,526 - SWAPO, Moses ||Garoëb (2014)
  4. 14,218 - SWAPO, Rundu Urban (2014)
  5. 14,065 - SWAPO, Rundu Urban (2019)

### Turnout Outliers

- **Minimum turnout:** 29.37% (Tobias Hainyeko, 2019)
- **Maximum turnout:** 138.47% (Windhoek East, 2019)
- **Records with turnout > 100%:** 180
- **Records with turnout < 40%:** 135

- **Mean turnout:** 67%

Geographic Range

- **Latitude:** -28.55° to -17.00° (valid for Namibia)
- **Longitude:** 13.24° to 24.63° (valid for Namibia)

## 7. Data Cleaning Recommendations

High Priority

- 1. Investigate turnout > 100%** (180 records)
  - Verify if same-day registration occurred
  - Cross-check with official electoral commission data
  - Consider creating adjusted turnout metric
- 2. Standardize party naming**
  - Presidential candidates appear as parties in 2019
  - Consider separate field for candidate vs party elections

Medium Priority

- 3. Examine zero-vote records** (48 records)
  - Verify legitimacy vs data entry errors
  - Document reasoning for keeping or removing
- 4. Handle high-vote outliers**
  - Validate top 5-10 highest vote records
  - Generally appear legitimate (SWAPO strongholds)

Low Priority

- 5. Remove or rename index column** (unnamed first column)
  - Currently just row numbers, not meaningful data
- 6. Consider creating derived variables**
  - Vote share percentage per constituency
  - Standardized turnout (capped at 100%)
  - Party competitiveness metrics

## 8. Transformation Requirements

Recommended Transformations

- 1. Turnout correction:** Create `turnout_adjusted = MIN(turnout_reg, 1.0)`
- 2. Vote share:** Add `vote_share = votes / total_votes * 100`

3. **Party categorization:** Separate presidential candidates from party records
4. **Region encoding:** Consider encoding for analysis (if needed for modeling)
5. **Date standardization:** Add explicit election date field if available

#### Part 4, Missing Values and Data Cleaning:

The dataset was already clean as there are no missing values and no duplicates. The data quality concern is that 180 voter records have voter turnout that exceeds 100%. The vote distribution is heavily skewed (median of 37 vs mean of 499), which is typical for electoral data where dominant parties receive substantially more votes. We will state this turnout anomaly clearly when analyzing and presenting these data. Therefore we do not need to perform Part 5 of this assignment, which is to remove duplicates.

#### Part 6, Two-Paragraph Reflection:

The exploration of the Namibia elections dataset showed the dataset to be a clean foundation with zero missing values and no duplicate records. This is a rare achievement in real-world electoral data. The summary statistics showed patterns of interest, particularly the extreme skewness in vote distribution where the median party receives only 37 votes while the mean is 499 votes. This analysis shows SWAPO's (SWAPO is the majority party) dominance, with some constituencies delivering over 17,000 votes to the party. The turnout statistics showed a mean of 67% participation. Our most concerning finding was that 180 records (3.5%) show mathematically impossible turnout rates (exceeding 100%, reaching as high as 138%) in Windhoek East. This anomaly, which might reflect same-day voter registration or temporal mismatches between registration snapshots and election day totals, doesn't invalidate the dataset but simply signals the need for adjusted metrics. We will not consider the turnout variable with as much weight as we previously planned. The geographic spread across 14 regions and 125 constituencies, combined with complete coordinate data, provides great spatial resolution for mapping electoral patterns. Given the spread and accuracy of these geographic data, and our conversations with course staff during office hours and in-class, we will plan to use the geoJSON package in our analysis so that we can geographically map these data.

Despite the turnout data quality concern, this dataset remains suitable to answer our questions about Namibian electoral politics. We can analyze which parties dominated specific regions, identify SWAPO strongholds versus competitive constituencies, and map the geographic distribution of party support. The 2014 versus 2019 comparison allows us to track party performance changes and possibly identify emerging parties (particularly visible in 2019's expanded party list). Questions about spatial patterns, such as whether northern regions vote differently than southern ones, are within our reach. Still, we need to consider the dataset's limitations as we do not have information about socioeconomic status of particular regions and other ethnographic information. Questions about how age, education, income, or ethnicity influenced voting behavior in the 2014 and 2019 Namibian elections are not answerable given these data.

## Milestone 4:

We met to discuss the following questions and responses:

1. Who is your audience? Come up with at least three options and pick one target audience.

Three potential audiences are first academics and political scientists interested in the matter, the ruling party of Namibia, and opposition parties of Namibia. We are picking the target audience of the opposition parties of Namibia so that they can learn more about regional party dynamics. The reason why we are targeting the audience

2. Describe your target audience in more detail. What do they know? What are their interests? What visualization literacy do they have? At what level of detail will you present information to them?

Our target audience knows the results of the elections, and they have an average visual literacy when it comes to data visualization. We have to be sure to clearly communicate trends and not insert any sort of bias into our projections so that our audience can take the visualizations at face-value. Their interests are in defeating the ruling party or taking away vote share in general from the ruling party. We will present information at a general scale and with a high level of detail so that if a politician is curious about a certain aspect of the data following reading our general overview, that politician will be able to look at any particular elements more closely.

3. What questions about your data will be interesting for your audience? Come up with a list of interesting questions that your audience may have about your data. The more, the better, but your team should come up with at least ten questions.

- How has SWAPO (the ruling party) lost vote share over time?
- How has vote share changed for opposition parties in the past six years and how can we use that to predict the next ten years?
- How does socioeconomic status correspond to political parties?
- Is there any evidence that particular political campaigns or efforts influenced vote share for a particular party?
- Why is SWAPO losing power and is there an opportunity for a lesser party to capitalize upon?
- How do ethnic groups and local leadership play into vote share, and how consistent is voting by region?
- Which constituencies were won by the smallest margins, and do these competitive areas cluster in certain regions?
- Do closer races tend to be associated with higher voter turnout?

- How does constituency size (registered voters specifically) relate to turnout, and are there anomalies where turnout is unexpectedly high or low?
  - Which regions exhibit the greatest party fragmentation, and how might that fragmentation create opportunities for opposition parties?
  - How do levels of fragmentation and competitiveness differ between urban and rural regions?
4. What data do you have or plan to obtain? Briefly describe the data you envision to use and the respective data types (categorical, ordinal, or quantitative) in your process book. It's OK if you are unsure about the data type for some attributes - you can simply describe them (e.g., geographic location).

We have data on the elections from 2014-2020 Namibian elections. The majority of these data are quantitative. We will attempt to pair these data with ethnic majority rulings and party vote share. We are going to analyze the ruling party's successes by region and then look at locations at which an opposition party has defeated the ruling party, and examine what made such locations different from others. Attributes interesting to us therefore include geographic location, political party, total votes, and election category.

## Catherine Mignone: Milestone 4

Catherine's Reflection:

I answered our question about where SWAPO lost vote share over time with my first two visualizations. Upon reflecting on those two visualizations, I decided to investigate voter turnout as a possible cause for the decline of the ruling party.

The questions addressed in these sketches are in line with our original questions. However, once I started looking through the data I became more interested in changes between the 2014 and 2019 elections because it quickly became very clear that the ruling party, SWAPO, had lost its dominant position between 2014 and 2019. While when we were developing our initial questions I assumed that I would be making a lot of bar graphs or linear representations (simply because these data are very quantitative as they concern vote shares and turnout percentages), when I began manually sketching out my first impressions of the dataset I became most interested in the dataset's categorical variables. Since so much of the dataset concerns regional information, I decided to attempt to visualize that information on an actual map so that an unfamiliar observer could understand any regional clustering without having to imagine where these regions might be. As I developed the maps, I realized that voter turnout might be very important especially because the ruling party experienced an upset between 2014 and 2019. These initial visualizations made me very excited about looking at these data in more detail, and I am especially excited for the most recent 2024 election data to come out!

Catherine's Visualizations:

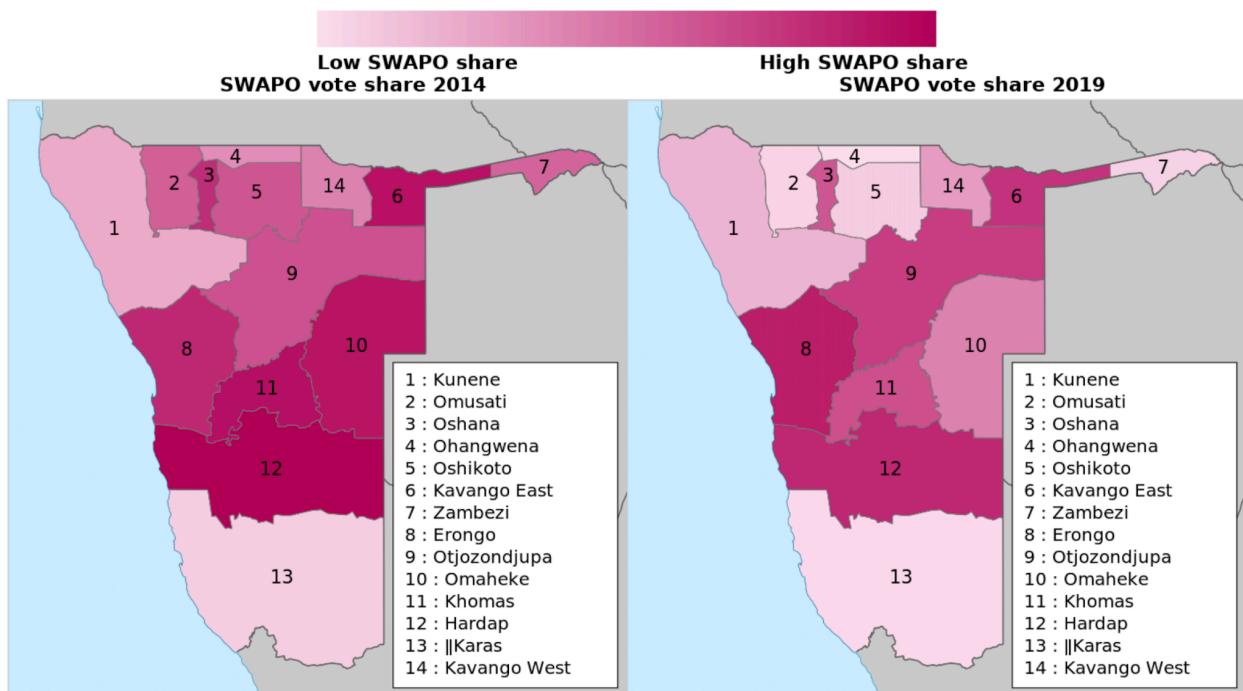
I used the dataset [linked](#) here.

I also extracted data from the government website [linked](#) here.

I used Claude and Chat GPT to help generate initial visualizations for this project. I hand-drew my goals for each visualization, and then created the drafts below digitally with Chat GPT.

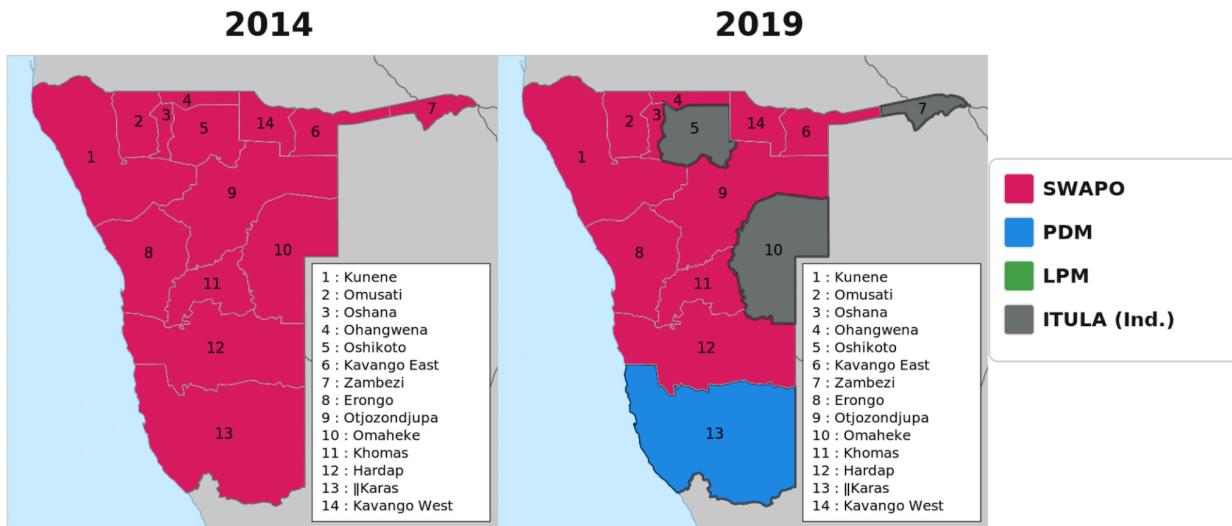
#### Visualization 1: SWAPO Vote Share, 2014 to 2019 Election Cycle Comparison

I used this [SVG file](#) for the visualization. I modified the file to shade the regions in, in accordance with SWAPO vote share.



#### Visualization 2: Winning Parties by Region, 2014 to 2019 Election Cycle Comparison

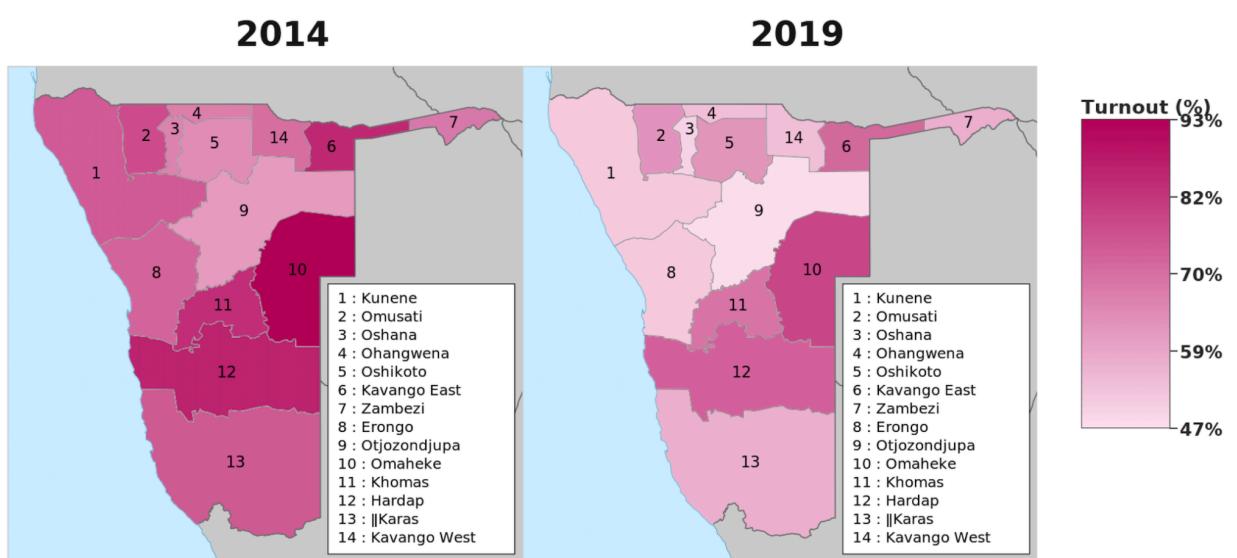
I used this [SVG file](#) for the visualization. I modified the file to shade the regions in, in accordance with vote share. I used Chat GPT for this visualization.



Visualization 3: Voter Turnout by Region, 2014 to 2019 Election Cycle Comparison

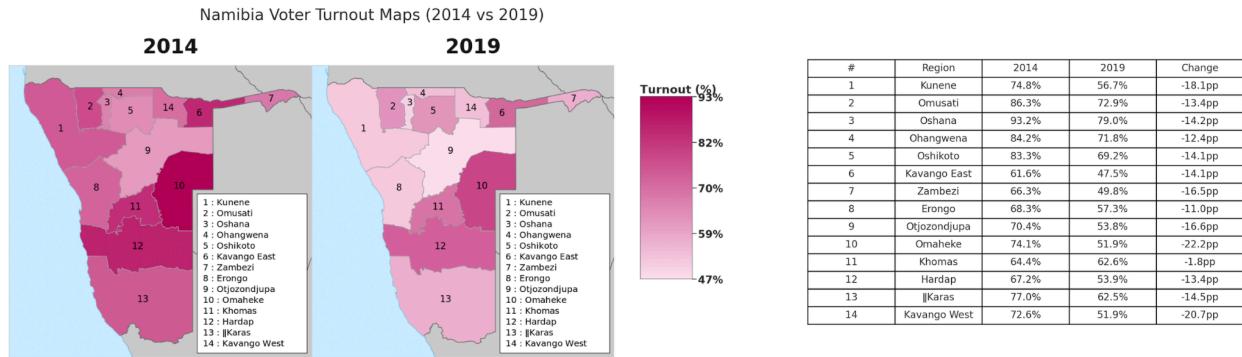
I used this [SVG file](#) for the visualization. I modified the file to shade the regions in, in accordance with vote share. I used Chat GPT for this visualization.

A. This is my first version of this visualization.



B. This is my second version of the visualization, where I added in a table to show the percent change in voter turnout so that a curious viewer could know more about the information

presented. Going forward I would like to workshop how to represent the numerical information directly onto the map instead of next to it.



## Matteo Wakeman: Milestone 4

### Matteo Wakeman Reflection:

I used the dataset [linked](#) here.

I also extracted data from the government website [linked](#) here.

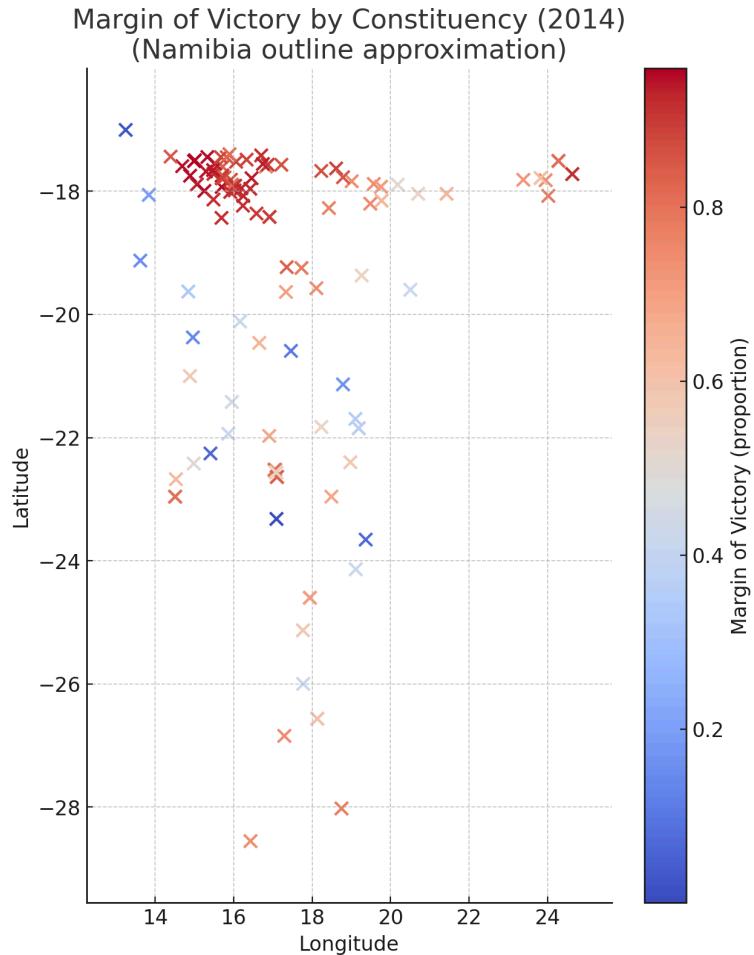
I used the dataset and relevant questions for each visualization, and then created the drafts below digitally with Chat GPT.

Overall, the questions I explored in my visualizations connect to our team's overall goals but take them in a slightly different direction. By focusing on competitiveness, turnout, and fragmentation (the last 5 questions we asked / tried to answer), I tried to highlight not only who won but the conditions behind those victories. The margin of victory map raised questions about which constituents are genuinely contestable and where sources could be focused. The turnout density maps encouraged me to think about whether constituency size influences participation and whether unusual patterns point to opportunities or weaknesses in mobilization. The fragmentation spiral shifted attention to how multiparty competition varies by region and where coalition-building might matter most. These questions differ slightly from the team's originals that centered more towards broad friends in vote share, but they complement them by offering a closer look at the electoral dynamics that create those trends in the first place. I think that these questions are especially valuable for our chosen audience since they highlight actionable insights rather than only descriptive ones.

### Visualization 1: Margin of Victory Heat Map

This visualization shows the margin between the winning party and the runner-up in each constituency during the 2014 election. Constituencies shaded in cooler tones represent close elections while red /

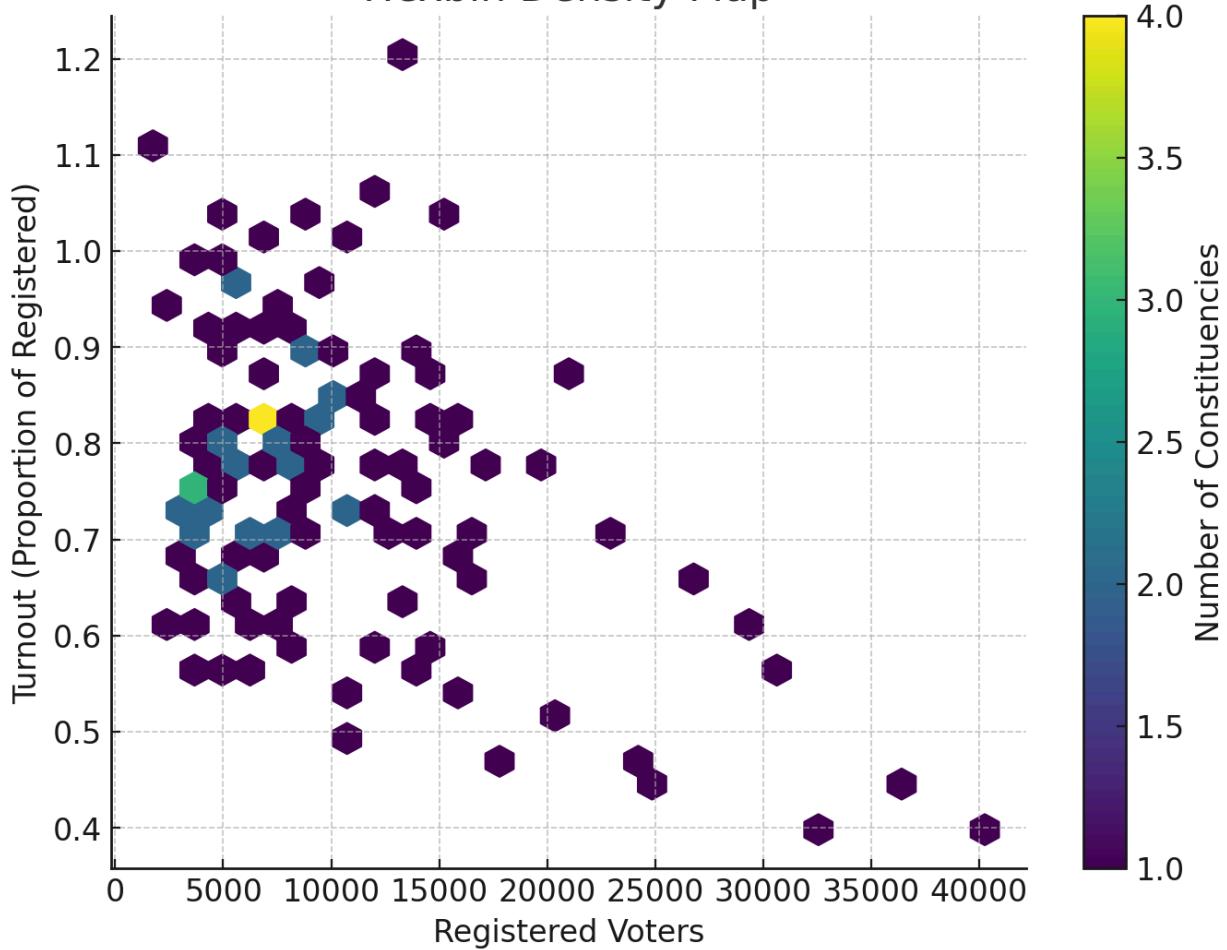
warmer tones mean landslide victories. This map highlights where elections were most competitive and where parties could allocate resources. I created it using the data from the dataset and ChatGPT.



### Visualization 2: Turnout vs Registered Voters Hexbin

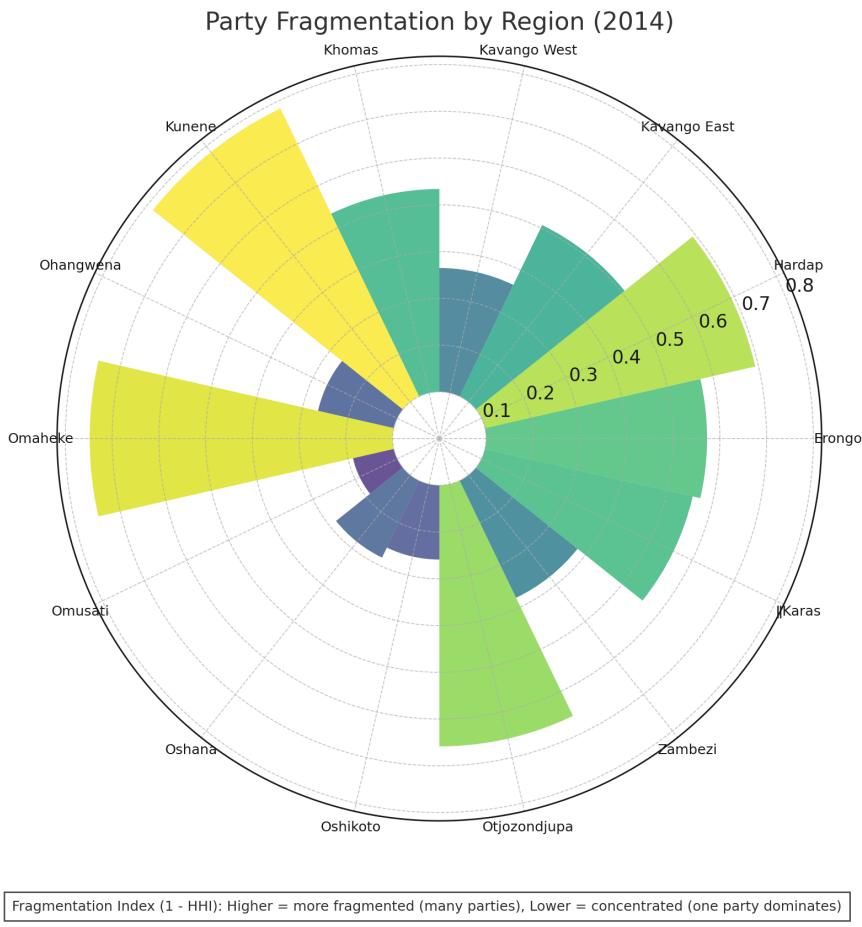
This visualization plots constituency size against voter turnout, with density shown through the hexagonal bins. The darker areas represent clusters of constituencies with similar patterns showing if there was systematically higher or lower turnout in voting. This helps flag any anomalies that may have occurred in the election. I created it using the data from the dataset and ChatGPT.

## Turnout vs Registered Voters (2014) Hexbin Density Map



### Visualization 3: Party Fragmentation Spiral

This chart measures fragmentation in each region. The circle layout makes it easy to compare it across regions. The larger the number, the larger the fragmentation by party (higher values indicate more fragmented systems). I created this visualization using the data and GPT to compute the HHI index and generate the visualization.



## Nathaniel Berkman: Milestone 4

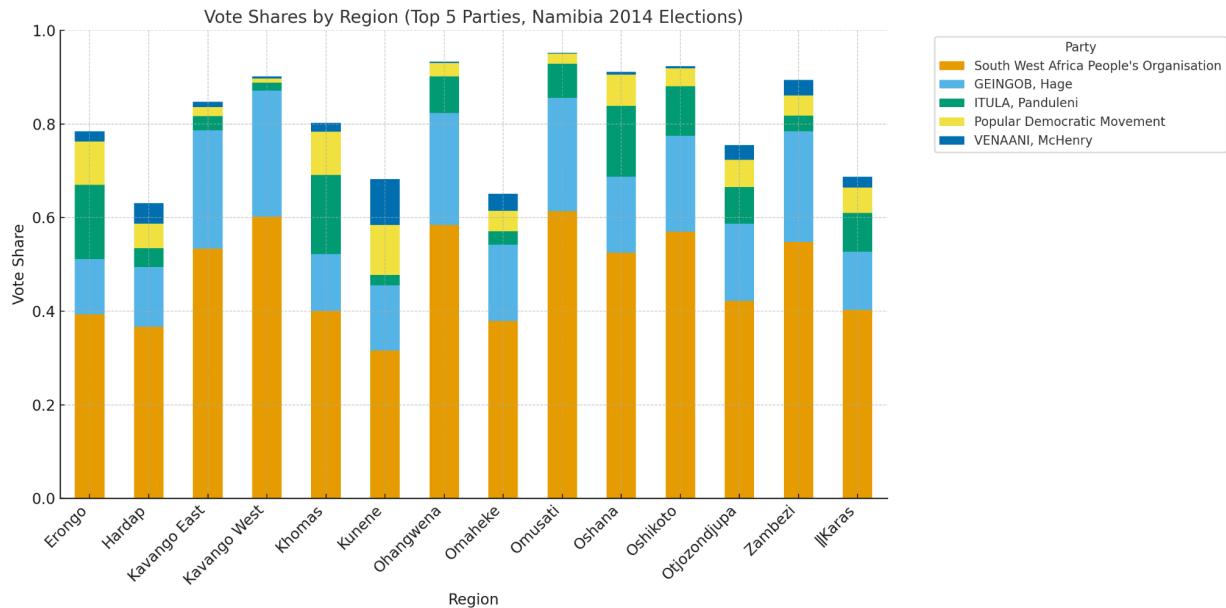
Nathaniel's reflection:

I answered the question regarding how regional party dynamics in Namibia's elections have shifted between 2014 and 2019. For my first two visualizations, I wanted to simply visualize each year, with vote share per region being the main data point displayed. However, when one looks within the bars, the breakdown per region is easily legible and adds a third dimension to the visualization. Visualizations 1 and 2 show the years 2014 and 2019 respectively, so in my third visualization, I wanted to combine the two and use order to show how breakdowns have changed over time. The question I am addressing differs slightly from the team's original questions as in my visualizations, I focus on how regional party dynamics have shifted between 2014 and 2019, instead of just the SWAPO's vote share. This question is essentially addressed when combining different aspects of our original questions, so it differs only

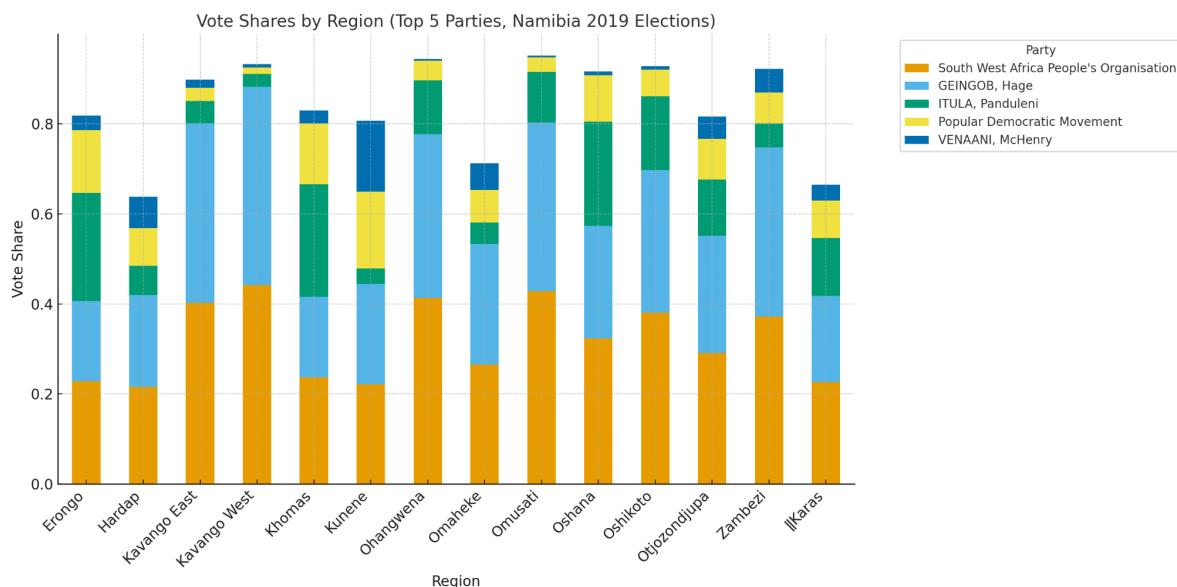
because I wanted to visualize aspects of different questions. Some questions are certainly better than others as our goal is to explain regional party dynamics to opposition parties, and the data requested in some questions is certainly less influential in the results of an election than that of other questions.

All 3 visualizations were created with the help of ChatGPT and excel. The data used from our data set was year, Name1, party, votes, and total\_votes and then this data was processed and normalized to create each of these charts.

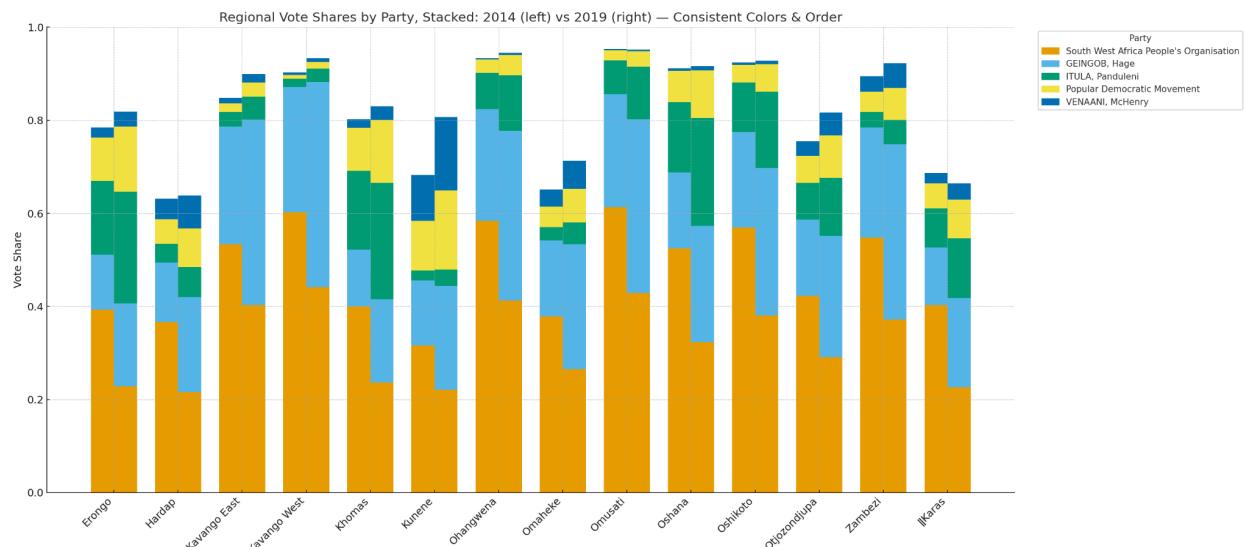
### Visualization 1:



### Visualization 2:



Visualization 3:



## Milestone 3: Project Proposal

**Title:** Namibian Election Dataset Analysis

**Team Members:** Catherine Mignone, Matteo Wakeman, Nathaniel Berkman.

**Team Name:** Vizualization Station!

### Background and Motivation

We chose this project because we are all interested in local politics, and Catherine's senior thesis topic is on diamond mines in Namibia. While election data likely will not play a role in her senior thesis, as a group we became interested in the country at large and found a convenient and large dataset that we think can answer some interesting research questions. Catherine did fieldwork in Namibia over the summer and found that ethnic groups play a large role in local government and corporate life in the capital city, Windhoek. Given these data, we are curious to see how such a political backdrop in the capital city plays out in more isolated regions of the country.

The three of us are really interested in political geography and we believe that this project will further our understanding and love for the topic. We have previously read more 'gimmicky' studies, for example about [local sports success and the election of incumbents](#), in relation to political geography. These studies are how we became interested in the topic, but quickly we discovered larger national geographical patterns. We want to look at these in Namibia because it

is a new setting for two of us which makes it an exciting academic experience, and for one of us it is a familiar topic but a new realm within it.

**Abstract:**

How does voter turnout, ethnic group, political party, and general geography influence election outcomes in Namibia?

The main goal of this project is to visualize the 2014 legislative elections in Namibia. The obtained dataset features information on constituencies, registered voters, turnout, total votes, and party-level results across regions. We want to explore patterns of voter turnout and political party performance both regionally and nationally in order to discover how support varies with geography. After analyzing regional tendencies, we will compare these themes with existing ethnic groups and alliances to see whether or not ethnic region plays a role in both political party support and voter turnout overall.

The visualization we first want to create is a choropleth map to visualize party dominance and turnout levels regionally. We are considering a variety of visualizations for showing party vote shares, and while we know it is simple we think a stacked bar chart would supplement the more complex choropleth mappings that we have planned. For our choropleth map, we want it to be an interactive website so that the viewer can roll their cursor over different regions and see more detailed information about each one. We would also appreciate advice over how to visualize ethnic and social groupings, as there are existing maps but we would like to combine those existing maps with our own data analysis and visualization. Further, we want to involve socioeconomic status by geography. Therefore our visualizations will take place on physical maps of Namibia, and our analysis will come across in how we differentiate between sub-regions and will be supplemented by additional simple graphic visualizations.

We would appreciate advice from course staff on how to create visualizations in accordance with geography. One idea we had was to take inspiration from the different subway map styles that we looked at in class, especially because exact physical proximity might not be necessary for understanding these data. We do understand that our impressions of these data will change once we begin performing simple analyses and simple visualizations of the dataset at hand prior to embarking on the more complex tasks we are aiming to perform.

We do believe that this project is thoughtful and unique, as we are looking at a combination of geography and political opinion in addition to population information. We will visualize these data in communication with each other, and the interactive maps that we aim to develop are an ambitious task for the semester. Still, this project is feasible within the constraints of a semester course project as we have obtained an existing dataset and have the necessary computer science

skills to execute what we have set out to do. We will have to learn these visualization techniques “on the fly” - but that is why we are in CS1710! We are really excited to take on this challenge.

**Data:** Catherine obtained data of Namibian election results in 2014 through her senior thesis research. The dataset is incredibly large, so we are going to spend time cleaning up the dataset. We link-share the data set [here](#).

Below is our plan for cleaning and pre-processing the dataset in the next few weeks. We will check off each task as we complete it in this process book.

- Assess data quality
  - Inconsistent names, mismatches, look at encoding problems.
  - Handle missing entries
- Validate content
  - Review the data manually because there may be spelling differences that we might have missed.
- Standardize:
  - Region names (there are 14 official regions)
  - Map hierarchy from polling station to region.
- Ensure all vote counts are non-negative integers.
- Verify vote shares sum to 100%.
- Check against registered voters and turnout.

## Team Agreement

We agree to support each other as a team and make the best project that we can! By following the below guidelines, we are confident that we can produce a great end result while enjoying the process of working together.

- ROLES AND RESPONSIBILITIES
  - Catherine: Finding the dataset + R
  - Matteo: Cleaning excel
  - Nathaniel: HTML/JS
- COMMUNICATION PREFERENCES
  - We have a group chat of the 3 of us and will use that to communicate with each other about the project.

- WORKING ARRANGEMENTS
  - We are going to meet weekly after class on Mondays for 30 minutes.
- WORK ALLOCATION
  - We've allocated this in our responsibilities, but we are going to report our progress on our tasks each Monday. This is why our meetings are short - it is for accountability.
  - We will also help each other out with our tasks and delegate as needed!
- ACCOUNTABILITY
  - Weekly meetings where we will report our completed tasks and work from the past week.

Signed,

Nathaniel Berkman, Catherine Mignone, Matteo Wakeman