Modellers and Story Tellers: Group 2 project

AUTHOR
Tendai Gwanzura

PUBLISHED
June 25, 2025

Hollywood coalition data cleaning and analysis.

This file contains the data cleaning and data analysis for Hollywood Food Coalition an organization aimimg to reduce food insecurity in Los Angeles.

.

Install and load libaries

install.packages("readxl") install.packages("openxlsx") install.packages("dplyr") install.packages("ggplot2") install.packages("janitor") install.packages("tidyr") install.packages("stringr") install.packages("tidytext") install.packages("topicmodels") install.packages("wordcloud") install.packages("RColorBrewer")

Load and Prepare the Data

Use read_excel to load data, after removing encryption in excel and rename new file to HFC_2025. For password openxlsx does not work despite having the password option.

```
HFC_2025 <- read_excel("HFC _2025.xlsx")
View(HFC_2025)

names(HFC_2025)
```

```
 [1] "Org ID"
 [2] "Orgs"
 [3] "Date"
 [4] "Organization Type"
 [5] "Location"
 [6] "Hours of Operation"
 [7] "On Site Contact"
 [8] "Metrics"
 [9] "Volunteer Run Organization"
[10] "Unique individuals served weekly"
[11] "Total served weekly"
[12] "Unique individuals served monthly"
[13] "Total served monthly"
[14] "Age range of population served"
[15] "Racial makeup of population served"
[16] "Percentage of low income population served"
[17] "(Under 18) Population Percentage"
```

```
[18] "(18-25) Population Percentage"
[19] "(26-40) Population Percentage"
[20] "(41-65) Population Percentage"
[21] "(65+) Population Percentage"
[22] "Population Languages"
[23] "(Male) Gender Make Up"
[24] "(Female) Gender Make Up"
[25] "(Non-Binary) Gender Make Up"
[26] "(Transgender) Gender Make Up"
[27] "Veterans served"
[28] "Previously Incarcerated"
[29] "LBGTQ+"
[30] "Unhoused"
[31] "Program Entry/Food restrictions"
[32] "Primary Demographics"
[33] "Primary Services"
[34] "Services Offered"
[35] "Referrals accepted"
[36] "Typical individual Seeking Service"
[37] "Outcome Goals"
[38] "Service Locations"
[39] "Recent Changes"
[40] "Meeting Recent Changes"
[41] "1 Year Growth Plans"
[42] "5 Year Growth Plans"
[43] "Holidays Recognized"
[44] "Other Schedule Notes"
[45] "Preferred Food Items"
[46] "Most to Least Desired"
[47] "Reason for Originally Partnering"
[48] "Reason for Continued Partnership"
[49] "Food Sources"
[50] "Food Intake Barriers"
[51] "(Hypothetical) Food Response without Hofoco"
[52] "(Hypothetical) Hofoco Food Delivery Option"
[53] "Recent Food Quality"
[54] "Food Quality"
[55] "Impact of Food Received"
[56] "Impact Story (1)"
[57] "Impact Story (2)"
[58] "Impact Story (3)"
[59] "Food Distro Process"
[60] "Food Distro Challenges"
[61] "Feedback"
[62] "Notes"
[63] "Future Check In Preference"
```

```r
# Total unique individuals served weekly
summary(HFC_2025$'Unique individuals served weekly')
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  20.00   85.25  185.00  315.23  312.50 3500.00       2
```

```
# Total individuals served monthly
summary(HFC_2025$'Unique individuals served monthly')
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   40.0   132.5   300.0   417.4   500.0  2000.0       3
```

The variable names contain spaces hence will need to be cleaned using the janitor package creating a new data set HFC_2025_clean.

```
 [1] "org_id"
 [2] "orgs"
 [3] "date"
 [4] "organization_type"
 [5] "location"
 [6] "hours_of_operation"
 [7] "on_site_contact"
 [8] "metrics"
 [9] "volunteer_run_organization"
[10] "unique_individuals_served_weekly"
[11] "total_served_weekly"
[12] "unique_individuals_served_monthly"
[13] "total_served_monthly"
[14] "age_range_of_population_served"
[15] "racial_makeup_of_population_served"
[16] "percentage_of_low_income_population_served"
[17] "under_18_population_percentage"
[18] "x18_25_population_percentage"
[19] "x26_40_population_percentage"
[20] "x41_65_population_percentage"
[21] "x65_population_percentage"
[22] "population_languages"
[23] "male_gender_make_up"
[24] "female_gender_make_up"
[25] "non_binary_gender_make_up"
[26] "transgender_gender_make_up"
[27] "veterans_served"
[28] "previously_incarcerated"
[29] "lbgtq"
[30] "unhoused"
[31] "program_entry_food_restrictions"
[32] "primary_demographics"
[33] "primary_services"
[34] "services_offered"
[35] "referrals_accepted"
[36] "typical_individual_seeking_service"
[37] "outcome_goals"
[38] "service_locations"
[39] "recent_changes"
```

```
[40] "meeting_recent_changes"
[41] "x1_year_growth_plans"
[42] "x5_year_growth_plans"
[43] "holidays_recognized"
[44] "other_schedule_notes"
[45] "preferred_food_items"
[46] "most_to_least_desired"
[47] "reason_for_originally_partnering"
[48] "reason_for_continued_partnership"
[49] "food_sources"
[50] "food_intake_barriers"
[51] "hypothetical_food_response_without_hofoco"
[52] "hypothetical_hofoco_food_delivery_option"
[53] "recent_food_quality"
[54] "food_quality"
[55] "impact_of_food_received"
[56] "impact_story_1"
[57] "impact_story_2"
[58] "impact_story_3"
[59] "food_distro_process"
[60] "food_distro_challenges"
[61] "feedback"
[62] "notes"
[63] "future_check_in_preference"
```

We can see a summary of the individuals served weekly and monthly.

```
summary(HFC_2025_clean$unique_individuals_served_weekly)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  20.00   85.25  185.00  315.23  312.50 3500.00       2
```

```
summary(HFC_2025_clean$total_served_weekly)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
    0.0   100.0   165.0   318.3   387.5  2100.0       2
```

```
summary(HFC_2025_clean$unique_individuals_served_monthly)
```

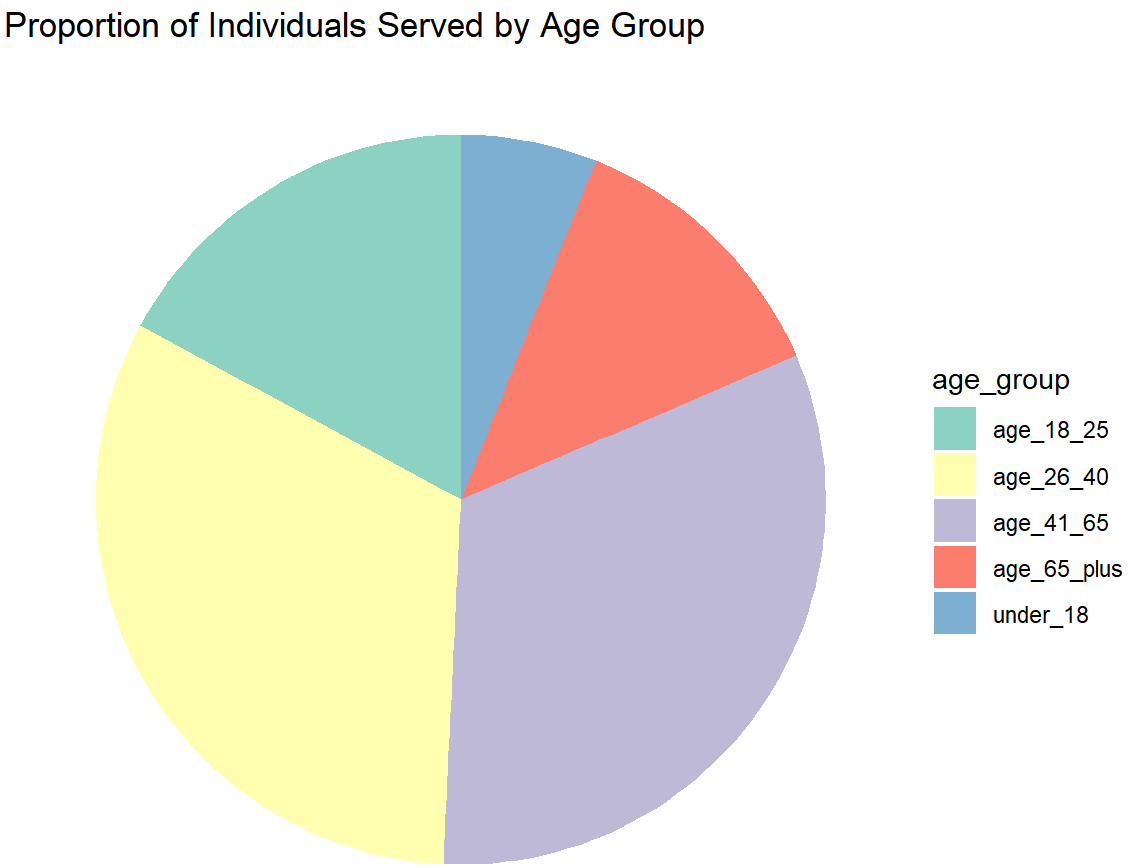
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   40.0   132.5   300.0   417.4   500.0  2000.0       3
```

```
summary(HFC_2025_clean$total_served_monthly)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
      0     376     625    1697    1550   20000       2
```

Age and gender distribution across all organizations

```
# A tibble: 1 × 5
  under_18 age_18_25 age_26_40 age_41_65 age_65_plus
     <dbl>     <dbl>     <dbl>     <dbl>       <dbl>
1   0.0609     0.172     0.325     0.325       0.126
```



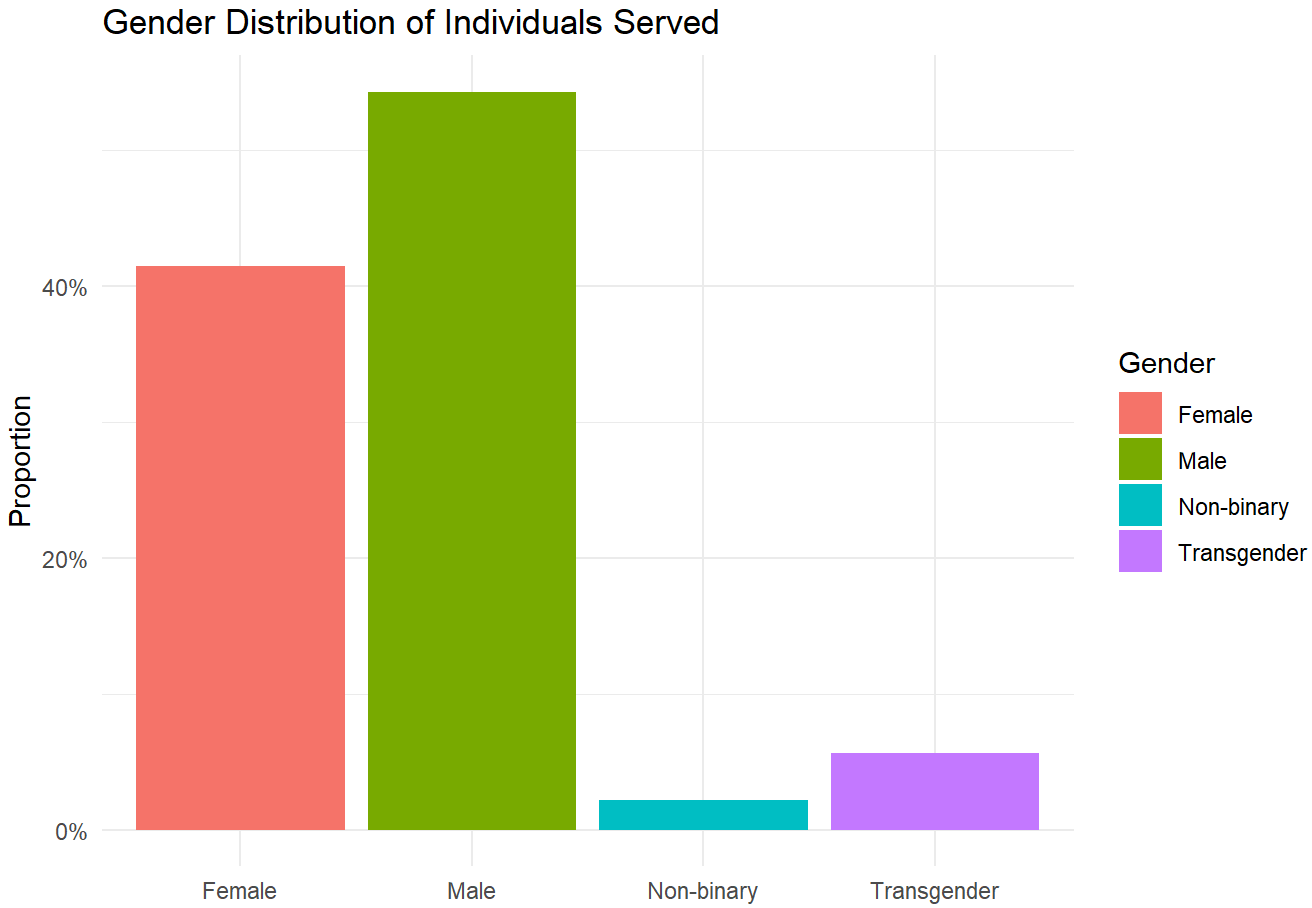
On average, 6.1% of individuals served across all organizations are under 18. Around 17.2% of individuals are young adults (18–25). About 32.5% are adults aged 26–40. About 32.5% are middle-aged to older adults (41–65). Around 12.6% are seniors, aged 65 and above.

```
# A tibble: 1 × 4
   male female non_binary transgender
  <dbl>  <dbl>      <dbl>       <dbl>
1 0.543  0.415      0.022      0.0565
```

About 54.3% of the individuals served are male.About 41.5% are female.About 2.2% are non-binary. About 5.6% identify as transgender (can overlap with other categories depending on how the data was reported). The following plot shows gender distribution across all partners.

```
# Using results from above to plot the gender
gender_data <- data.frame(
  Gender = c("Male", "Female", "Non-binary", "Transgender"),
  Proportion = c(0.5432432, 0.4145946, 0.022, 0.05645161)
)
```

```
ggplot(gender_data, aes(x = Gender, y = Proportion, fill = Gender)) +
  geom_bar(stat = "identity") +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1)) +
  labs(title = "Gender Distribution of Individuals Served",
       y = "Proportion",
       x = "") +
  theme_minimal()
```



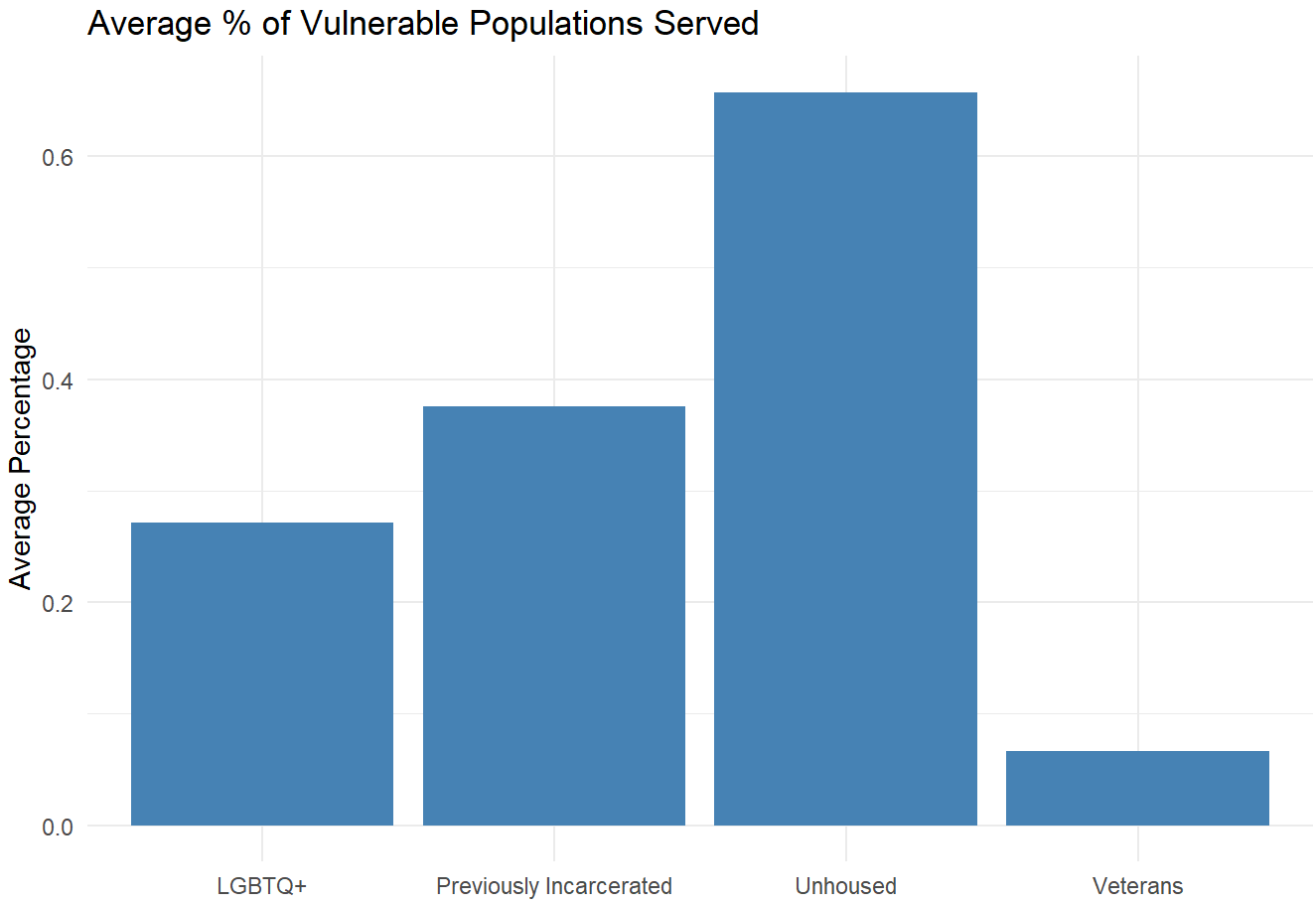
The following shows the distribution of the different populations in the dataset.

```
pop_focus <- HFC_2025_clean %>%
  summarise(
    unhoused = mean(as.numeric(unhoused), na.rm = TRUE),
    lgbtq = mean(as.numeric(lbgtq), na.rm = TRUE),
    veterans = mean(as.numeric(veterans_served), na.rm = TRUE),
    previously_incarcerated = mean(as.numeric(previously_incarcerated), na.rm = TRUE)
  )
pop_focus
```

```
# A tibble: 1 × 4
  unhoused lgbtq veterans previously_incarcerated
     <dbl> <dbl>    <dbl>                   <dbl>
1    0.657 0.272   0.0667                   0.376
```

```r
pop_data <- data.frame(
  Group = c("Unhoused", "LGBTQ+", "Veterans", "Previously Incarcerated"),
  Percentage = c(
    pop_focus$unhoused,
    pop_focus$lgbtq,
    pop_focus$veterans,
    pop_focus$previously_incarcerated
  )
)

ggplot(pop_data, aes(x = Group, y = Percentage)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(
    title = "Average % of Vulnerable Populations Served",
    y = "Average Percentage", x = ""
  ) +
  theme_minimal()
```



Cleaning and plottin race

Race contains mixed data that is string and numeric we might need to separate it using the following code.

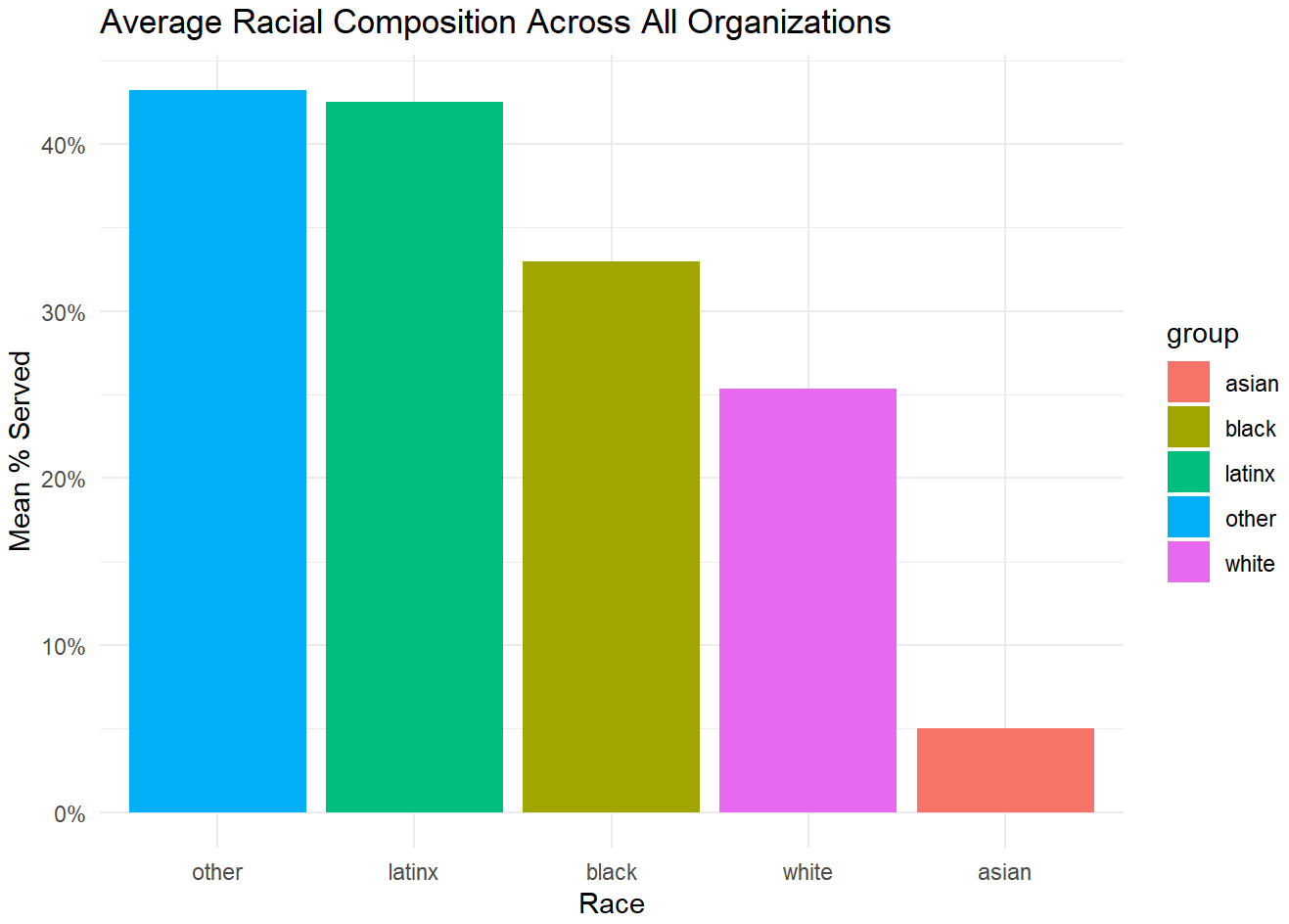
```
 [1] NA
 [2] "50% Black 50% Brown"
 [3] "Asian Diaspora/ Pacific Islander"
 [4] "Latino 50%, White 30%, Black 10%, Other 10%"
 [5] "white 60%, Latino 25%,. Black 10%, Other 5%"
 [6] "White 50%, Latino 30%, Black 20%"
 [7] "White 40%, Black 25%, Latino 20%,  Other 10%, Asian 5%"
 [8] "Latinos, Whites, Blacks"
 [9] "Black 42%, 35% Latino, 17% White, 15% Other"
[10] "60% Filipino, 20% Latino, 10% White, 10% Balck"
[11] "Latino 60%, Black 30%, Other 10%"
[12] "Hispanic 44%, Black 35 %, White 21%,"
[13] "40% Latino"
[14] "Latino 50%, White 25%, Black 24%, Asian 1%"
[15] "Hispanic 49%, Latino 49%, White 2%"
[16] "Latino 32%, Black 31%, White 29%, Other 8%"
[17] "Latino 55%, Armenian 35%, Arab 10%"
[18] "Latino70%, Black 20%, White/Other 10%"
[19] "Latino 40%. Black 30%, Other 30%"
[20] "Black, Latino, and Asians"
[21] "Latino 75%, Black 10%, API 5%"
[22] "Latino 75%, Asian 25%,"
[23] "White 35%, Multi 32%, Latino 16%, Black 12%, Other 5%"
[24] "Armenian/Russian 50%, Latino 40%, Other 10%"
[25] "Korean 25%, Black 25%, Latino 25%, Filipino 5%, White 10%, Other 10%"
[26] "88% BIPOC"
[27] "Latino 86%, Other 14%"
[28] "50% White, 30% Black, 5% Latino, 5% Asian, 10% Other"
[29] "Black 60%, Latino 30%, White 10%"
[30] "White 50%, Latino 35%, Black 15%,"
[31] "42.5% Brown, 25% Black, 15% White"
[32] "No one race more than another"
[33] "Latino 20%, Asian 80%"
[34] "100% Latino"
[35] "50% Black, 40% Latino, 10% White"
[36] "50% White, 50% Hispanic"
[37] "Latino 75%, White 10%, Black 5%, API 5%"
[38] "Hispanic, Colombian, Salvadorian, Venezulian"
```

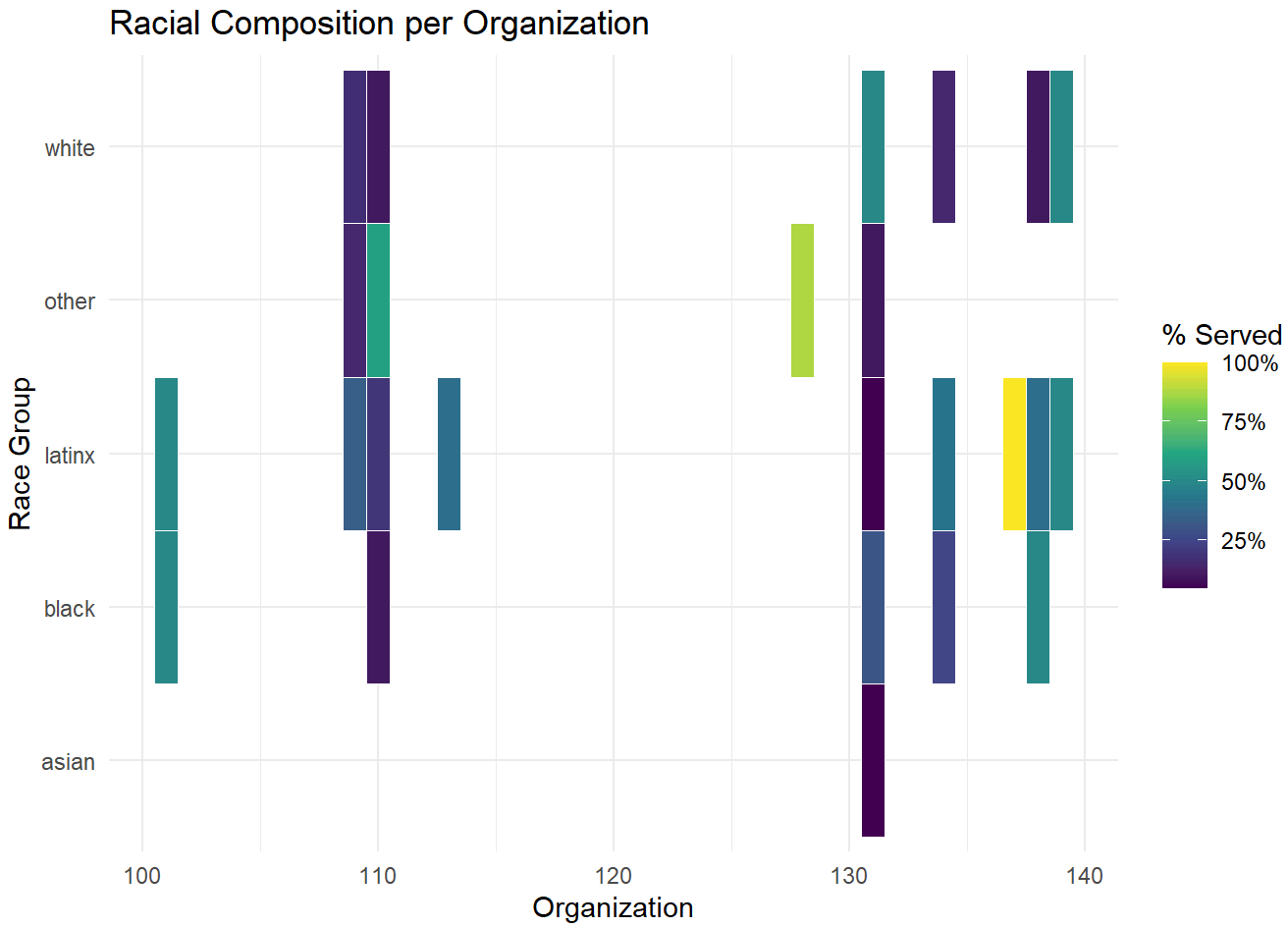
Important to note not all organizations include the percent of racial makeup they just list the races. The following plot assesses data which contains % makeup of the 38 out of the total 42 organizations without missing data. Filipino, Russian, Armenian and other country specific were set to other for now hence the large percent of other.

```
race_summary <- race_long %>%
  group_by(group) %>%
  summarise(mean_pct = mean(value, na.rm = TRUE))

ggplot(race_summary, aes(x = reorder(group, -mean_pct), y = mean_pct, fill = group)) +
  geom_bar(stat = "identity") +
```

```
  scale_y_continuous(labels = scales::percent_format()) +
  labs(title = "Average Racial Composition Across All Organizations",
       y = "Mean % Served", x = "Race") +
  theme_minimal()
```



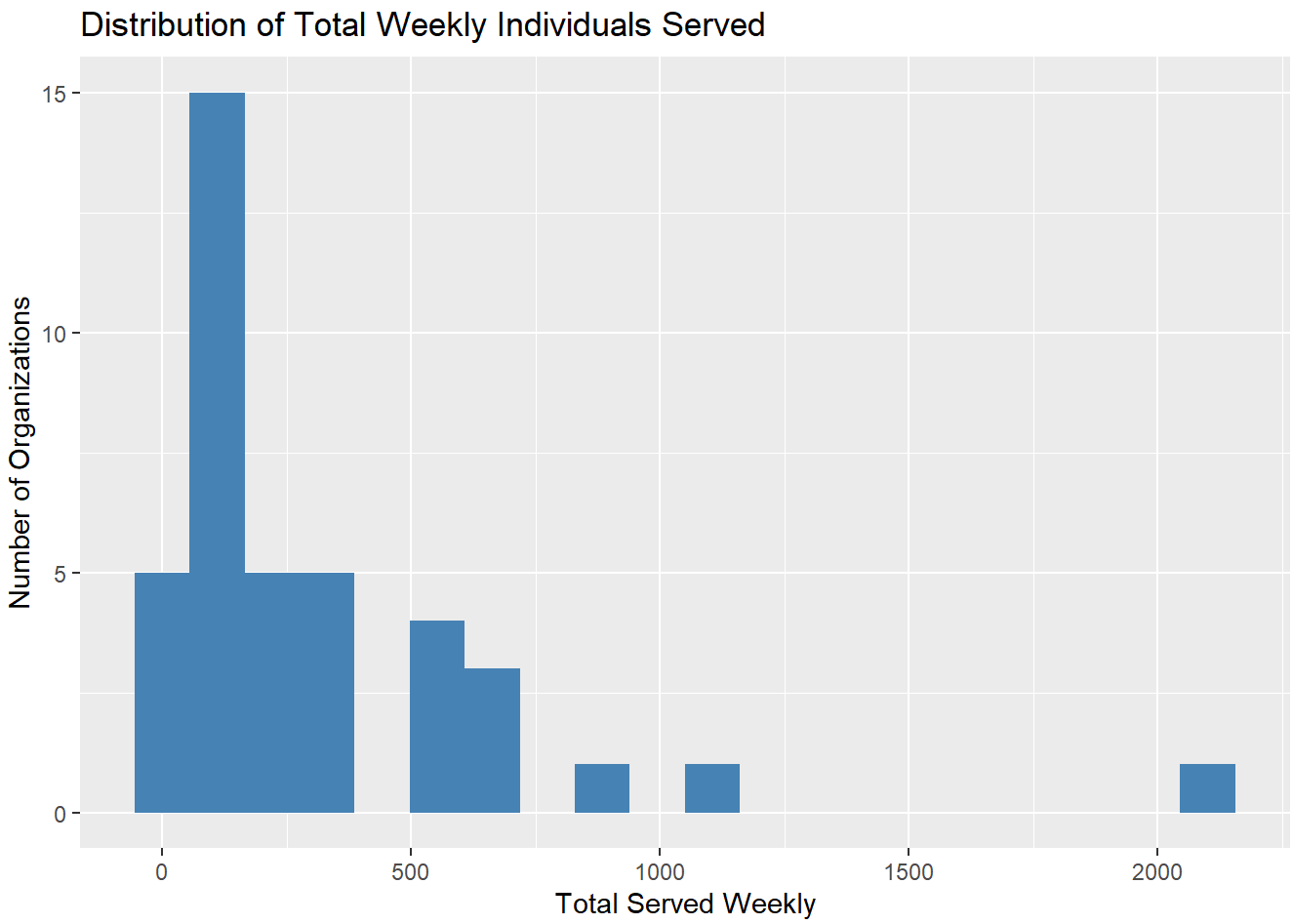


This visualization highlights how racial representation varies by organization. While Latinx and Black communities are consistently represented, some organizations show a near-exclusive service to one group. This racial composition is important for understanding equity in food access and can inform tailored programming and outreach. Notably, Asian and 'Other' populations appear less frequently, which could suggest either demographic realities or potential service gaps. Also not all organizations iclude percent of people affected.

Plotting the reach and scale

```
 unique_individuals_served_weekly total_served_weekly
 Min.   :  20.00                  Min.   :   0.0
 1st Qu.:  85.25                  1st Qu.: 100.0
 Median : 185.00                  Median : 165.0
 Mean   : 315.23                  Mean   : 318.3
 3rd Qu.: 312.50                  3rd Qu.: 387.5
 Max.   :3500.00                  Max.   :2100.0
 NA's   :2                        NA's   :2
 unique_individuals_served_monthly total_served_monthly
 Min.   :  40.0                     Min.   :    0
 1st Qu.: 132.5                     1st Qu.:  376
 Median : 300.0                     Median :  625
 Mean   : 417.4                     Mean   : 1697
 3rd Qu.: 500.0                     3rd Qu.: 1550
```

```
Max.   :2000.0                  Max.   :20000
NA's   :3                       NA's   :2
```



Qualitative Impact

Utilising word cloud to view impact.The default stopwords("en") removes general English filler words (like "the", "is", "it"). But our text still contains domain-irrelevant or generic words like: guys, week, kids, know, one, used, get, great, makes, just, able, come, etc.

The most frequent words include:

“food”, “people”, “give”, “quality”, “produce”, “relief”, “nutritious”, “kitchen”, “residents”, “parents”, “love”

The word ‘food’ dominates the stories — underscoring its centrality to service impact.

‘Relief’, ‘quality’, and ‘produce’ point to the value and trust recipients associate with the program.

References to ‘residents’, ‘parents’, and ‘love’ highlight emotional and familial dimensions of food insecurity.

We can go further and create themes based on the words.

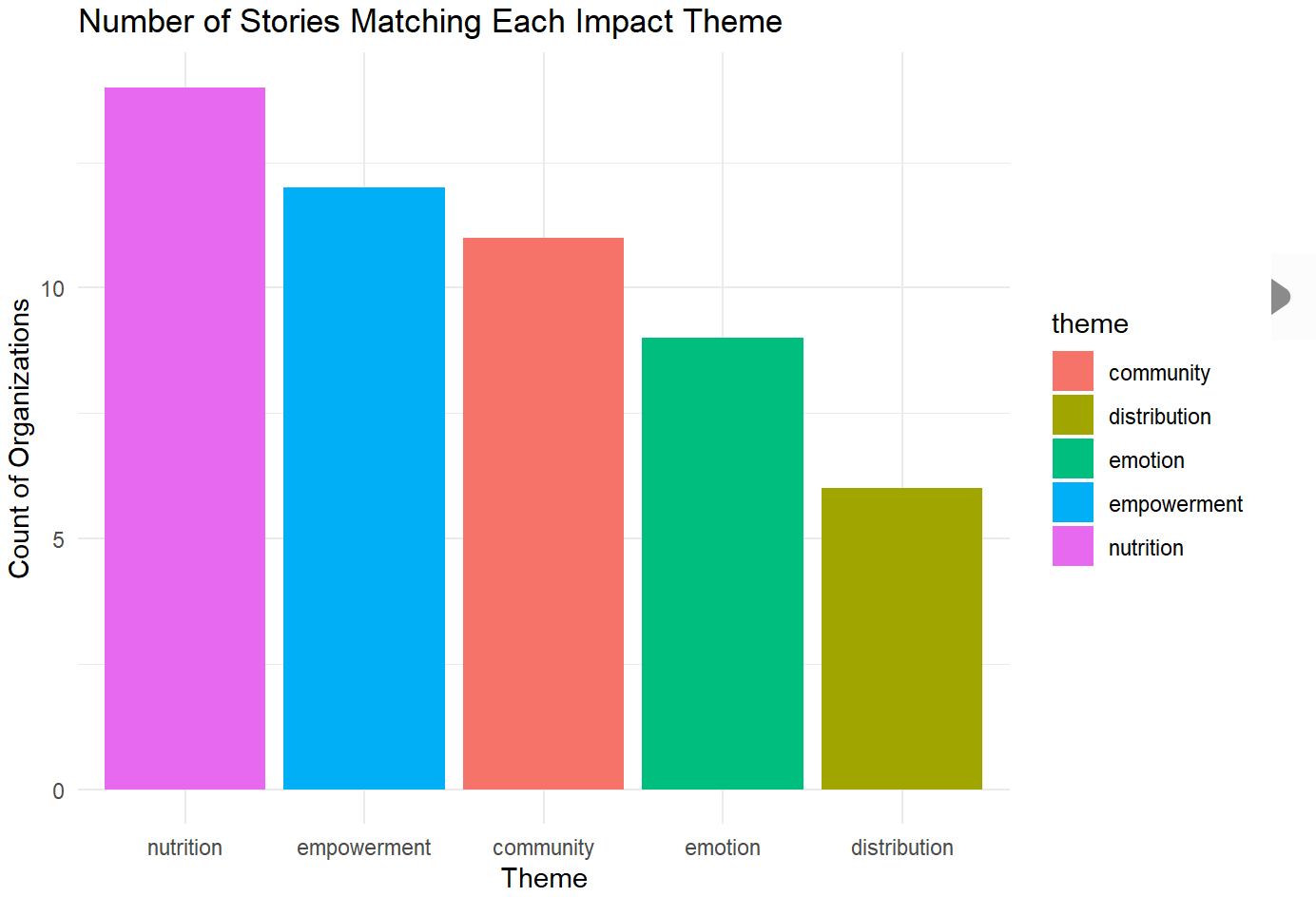
For this I chose nutrition, community, empowerment, distribution, emotion.

```
theme_summary <- theme_tags %>%
  select(nutrition, community, empowerment, distribution, emotion) %>%
  summarise(across(everything(), ~sum(.)))

theme_summary_long <- pivot_longer(theme_summary, everything(), names_to = "theme", values_to = "

ggplot(theme_summary_long, aes(x = reorder(theme, -count), y = count, fill = theme)) +
  geom_col() +
  labs(title = "Number of Stories Matching Each Impact Theme",
```

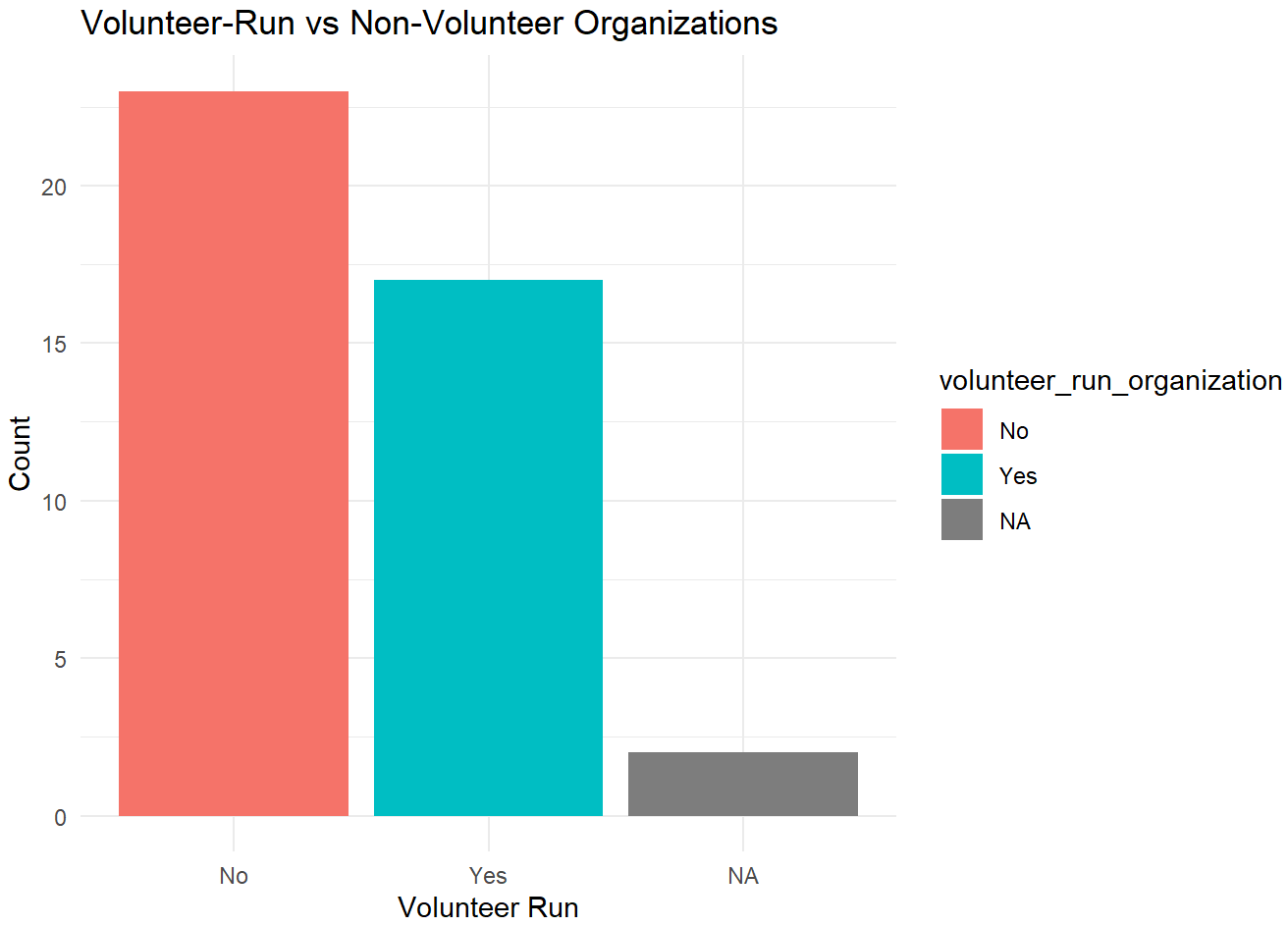
```
      x = "Theme", y = "Count of Organizations") +
  theme_minimal()
```



Note we can go further and do topic modelling but will leave it at themes for now.

Organization value and resilience

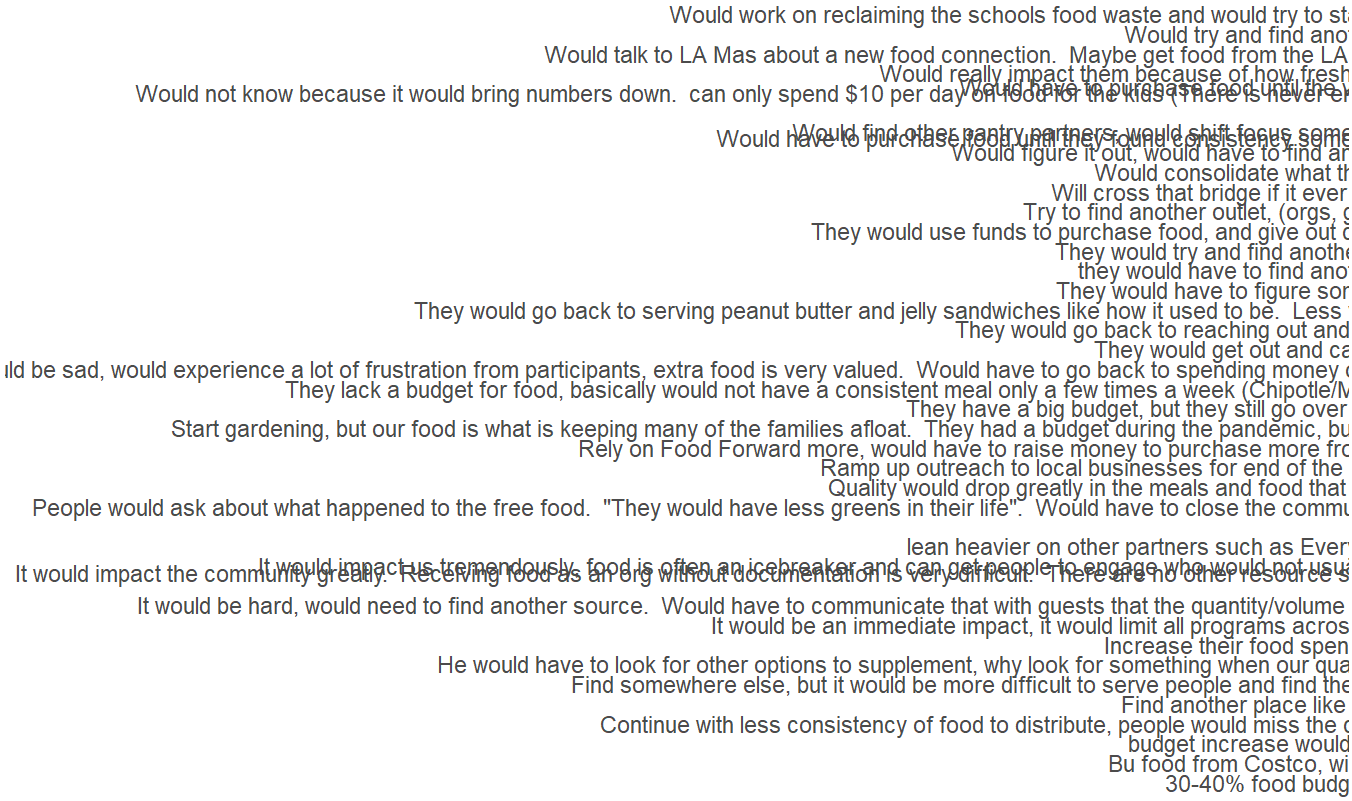
```
HFC_2025_clean %>%
  count(volunteer_run_organization) %>%
  ggplot(aes(x = volunteer_run_organization, y = n, fill = volunteer_run_organization)) +
  geom_bar(stat = "identity") +
  labs(title = "Volunteer-Run vs Non-Volunteer Organizations",
       x = "Volunteer Run", y = "Count") +
  theme_minimal()
```



Nearly half of the organizations that HoFoCo works with with are volunteer-run, highlighting a strong reliance on community-driven labor. However, a slightly greater number operate with non-volunteer staffing structures. This suggests a mix of grassroots and institutional capacity within the partner network. The presence of missing values also points to potential opportunities to improve data collection clarity or completeness."

Text exploration

Importance of HoFoCo

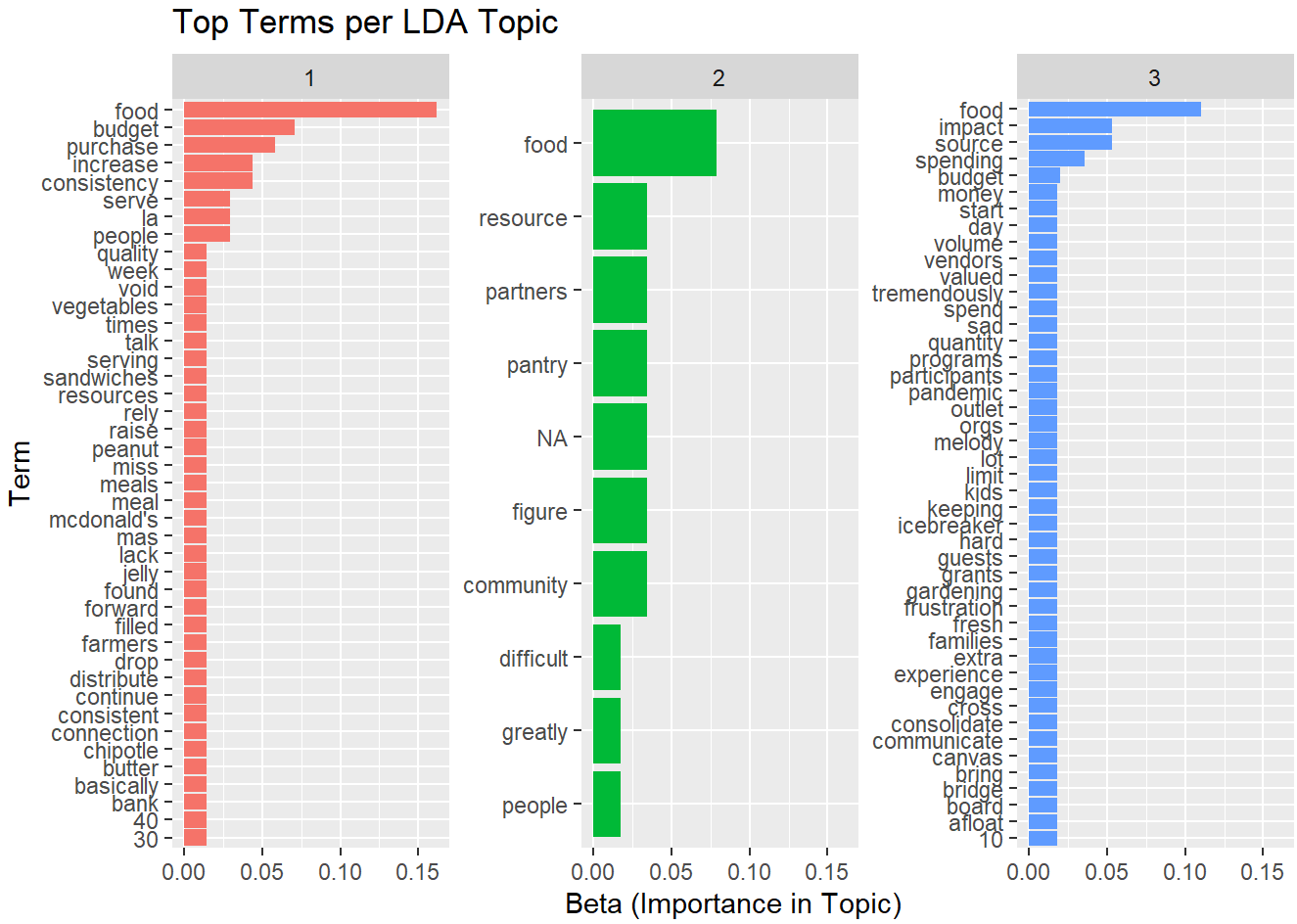



```
# A tibble: 97 × 3
   topic term          beta
   <int> <chr>        <dbl>
 1     1 food        0.162
 2     1 budget      0.0710
 3     1 purchase    0.0579
 4     1 increase    0.0434
 5     1 consistency 0.0434
 6     1 serve       0.0290
 7     1 la          0.0290
 8     1 people      0.0290
 9     1 quality     0.0145
10     1 farmers     0.0145
# ℹ 87 more rows
```

Top 10 keywords: food, budget, purchase, increase, consistency, serve, quality, farmers based on probability values. Now to view all 97 in topics.

```
top_terms <- topics %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)
```

```
ggplot(top_terms, aes(x = reorder_within(term, beta, topic), y = beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free_y") +
  coord_flip() +
  scale_x_reordered() +
  labs(title = "Top Terms per LDA Topic",
       x = "Term", y = "Beta (Importance in Topic)")
```



Using Latent Dirichlet Allocation (LDA), we identified 3 major thematic clusters from open-text responses in the “hypothetical_food_response_without_hofoco” question:

Topic 1: Food Budgeting and Procurement Top terms: food, budget, purchase, consistency, increase, serve, resources, raise, distribute, continue This topic centers around how organizations plan for and manage food procurement, budgeting constraints, and ensuring consistent distribution.

We can label topic as “sustaining Food Supply through Budgeting and Purchasing”

Topic 2: Partner Reliance and Community Coordination Top terms: food, resource, partners, pantry, community, figure, difficult, people This topic shows organizations’ dependence on partner networks, food pantries, and community coordination to address gaps. We can label it community Collaboration and External Resource Navigation

Topic 3: Urgency, Impact, and Emotional Strain Top terms: impact, source, spending, quantity, volume, tremendously, programs, participants, frustration, families, communicate This topic addresses the the emotional and operational impact of food insecurity and limited resources — particularly on participants and families. We can label it emotional and Operational Impact of Insecurity on Families

Note we can do more than 3 themes but for the purpose of presentation I just selected 3.

Lets try to run a regression next- work in progress