

# Tequila meets Technology

A new look at Tequila

Robert Gravelle

2023-11-28

## Contents

<b>Table of Contents</b>	<b>1</b>
<b>Introduction</b>	<b>3</b>
<i>A brief history of Tequila</i> . . . . .	3
<i>Technology helping traditions</i> . . . . .	3
<i>Loading and Exploring Data</i> . . . . .	4
Data Summary . . . . .	4
<b>Exploratory Data Analysis</b>	<b>5</b>
Tequila Exploratory Analysis . . . . .	5
Regions . . . . .	5
Cooking Types . . . . .	5
NOM Top ten producers . . . . .	6
Count by Extraction methods . . . . .	6
Tequila Production by Distillation Method . . . . .	7
Count by Still Methods . . . . .	7
Count by Water Source . . . . .	7
Count by ABV . . . . .	8
<b>Data Analysis and Modeling</b>	<b>9</b>
Ratings for overall Regions . . . . .	9
Ratings from Community and Panel raters . . . . .	10
Density plot ratings for the panel and community . . . . .	11
Including Plots . . . . .	12

## Table of Contents

# Contents

# Introduction

## *A brief history of Tequila*

Once upon a time in the sun-drenched lands of Mexico, there existed an ancient spirit with roots deeply embedded in the soil and traditions of the indigenous peoples. Long before the arrival of the Spanish conquistadors, the agave plant was revered for its versatile uses. The native people fermented its sap to craft a magical elixir known as “pulque,” a beverage that would set the stage for the birth of a legendary spirit.

As the Spanish adventurers, led by the intrepid Hernán Cortés, set foot on Mexican soil in the 16th century, they brought with them the knowledge of distillation. The alchemy of turning humble agave into a distilled spirit began, giving birth to what we now know as tequila.

The stage for tequila’s grand performance was set in the picturesque town of Tequila and the surrounding regions of Jalisco. Here, the volcanic soil and ideal climate nurtured the blue agave plant, which flourished under the watchful gaze of the Jalisco sun.

In 1758, a pioneer named José Antonio Cuervo established the first licensed tequila distillery, marking a pivotal moment in the spirit’s history. The golden elixir, aged in oak barrels, soon captured the hearts and palates of those who indulged in its magic.

The Mexican government, recognizing the importance of preserving tequila’s authenticity, introduced regulations in the 20th century. In 1974, the Denomination of Origin (DO) was established, defining the geographical boundaries where true tequila could be crafted. These regulations specified the types of agave permitted, with the blue agave taking center stage in the tequila tale.

As the years passed, tequila transcended borders, becoming an international sensation synonymous with fiestas and celebrations. The distinct categories of Blanco, Reposado, Añejo, and Extra Añejo added layers to the story, each contributing its unique flavor to the narrative.

Tequila became not just a drink but a cultural symbol, woven into the fabric of Mexican identity. It found its way into the hands of storytellers, artists, and revelers worldwide, leaving an indelible mark on the global spirits scene.

And so, the legend of tequila continues to unfold—a tale of agave fields whispering in the breeze, of distilleries crafting liquid gold, and of glasses raised in celebration. It is a story that intertwines the ancient traditions of Mexico with the modern rhythms of a world captivated by the spirit of tequila. And with each sip, the journey through time and taste goes on, inviting all to savor the magic born from the heart of the agave plant and the spirit of a land called Mexico.

## *Technology helping traditions*

Tequila aficionados are always on the lookout for novel and exceptional spirits, yet the hesitation to invest in a bottle without certainty of appreciation remains a common concern. Enter technology—a transformative force not only for recording and exploring ratings but also for predicting them. The key lies in deciphering the elements that constitute a stellar tequila, encompassing Agave regions, tequila categories, water sources, distillation techniques, production facilities, and the skilled artisans behind each creation.

Anticipating ratings necessitates a comprehensive understanding of current ratings and trends spanning several years. Embracing technology allows us to chart future trends and ratings for tequilas, enabling consumers to make informed decisions about unfamiliar spirits. It’s a digital guide, providing more than just personal preferences—a data-driven compass predicting the trajectory of tequila experiences.

This innovative approach not only empowers consumers to navigate uncharted tequila territories with confidence but also exemplifies the harmonious synergy between technology and the beverage industry. As we witness this convergence, technology becomes a beacon for enthusiasts and connoisseurs, offering insightful solutions that elevate the exploration of tequila to new heights.

## ***Loading and Exploring Data***

This project revolves around loading insights from three key files: `fulldataman.csv`, `weather.csv`, and `ybyratings.csv`. Picture these files as the culmination of an intricate ballet between data scraping from various websites and the continuous processing by diverse programs. The beauty lies in the fact that these files are not static; they're a reflection of the most recent tequila landscape.

As we ride the wave of data evolution, the files undergo meticulous preprocessing, acting as guardians of data integrity. This ensures that with every update the analysis remains robust, adapting seamlessly to the ever-shifting dynamics of tequila information.

One of the web scraping algorithms I wrote is contained below.

```
library(dplyr) library(rvest)

brands=NULL for(i in 1:58) { url=paste("https://tequilamatchmaker.com/brands?page=",i,sep="") wc <-
read_html(url) urls= wc %>% html_nodes('tr.clickable-row') %>% html_attr('data-href') urls=paste("https:
//tequilamatchmaker.com",urls,sep="")

brands=c(brands,urls) print(i)
}

brands=unique(brands) write.csv(unique(brands),"brands.csv")

allurls=NULL ua <- "Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:106.0) Gecko/20100101 Firefox/106.0"
for(i in brands) {
ii <- GET(i, user_agent(ua))

wc <- read_html(ii) urls= wc %>% html_nodes('a.product-list__item-link.material-shadow.material-shadow-
hover') %>% html_attr('href') urls=paste("https://tequilamatchmaker.com",urls,sep="") allurls=c(al-
lurls,urls) print(i) Sys.sleep(.5) } write.csv(unique(allurls),"allurls.csv")
```

## **Data Summary**

In the symphony of data, we have a trio of CSV files to elevate the tequila analysis. These live files undergo a preprocessing, into the file `fulldataman.csv`. This data source encapsulates the essence of Tequila, featuring names, NOM, distillation processes, ratings, flavors, aromas, and more. Imagine it as a distilled blend of information, meticulously collected from various online sources and expertly merged into this data haven.

The second file is the `weather.csv`, a meteorological data source of tequila and agave growth regions. Location, sunshine hours, humidity, temperature – this file paints a vivid meteorological portrait sourced from multiple online platforms.

Completing the third file is `ybyratings.csv`, a collection of summarized tequila ratings spanning the years 2020, 2021, and 2022. This file acts as a compass for trend analysis, meticulously filtering ratings based on the review dates. Together, these files form a grand ensemble, ready to serenade your tequila exploration.

# Exploratory Data Analysis

First and foremost, we must amass a diverse set of data essential for exploring the myriad facets of Tequila. This includes insights into Tequila by regions, cooking methods, NOM production, extraction methods, distillation methods, and ABV.

## Tequila Exploratory Analysis

### Regions

Now let's embark on an exploration of the regions where Tequila is produced.

Region	Total
Jalisco (Los Valles)	2498
Jalisco (Los Altos Southern)	1442
Jalisco (Central)	479
Jalisco (Cinega)	389
Guanajuato	44
Jalisco (Southeast)	33
Michoacn	29
Jalisco (Southern)	21
Tamaulpas	13
Jalisco (South Coast)	6
Jalisco (Sierra de Amula)	3
Jalisco (Sierra Occidental)	2
Jalisco (Los Altos)	1
Jalisco (North Coast)	1

### Cooking Types

Now let's embark on an exploration of the cooking methods for Tequila

Cooking	Total
Stone/Brick Ovens	2324
Autoclave (high pressure)	1593
Autoclave (low pressure)	179
Stone/Brick Ovens Autoclave (high pressure)	134
Acid-Thermal Hydrolysis	108
Autoclave (post diffuser)	77
Boiling Tank	13
Autoclave (high pressure) Stone/Brick Ovens	12
Stone/Brick Ovens Autoclave (low pressure)	9
Stone/Brick Ovens Boiling Tank	9
Earthen Pit	5
Stone/Brick Ovens Autoclave (post diffuser)	5
Steel Oven (Direct Flame)	4
Earthen Pit Autoclave (high pressure)	2
Stone/Brick Ovens Diffuser	2
Diffuser	1

## NOM Top ten producers

Now let's embark on an exploration of the top ten NOM

Table 1: Count by NOM (Top 10)

NOM	Total
1438	249
1459	226
1137	223
1499	160
1480	134
1466	130
1477	115
1173	109
1479	109
1414	99

## Count by Extraction methods

Now let's see the Extraction method of Tequila by method

Table 2: Count by Extraction Methods

Extraction	Total
Roller Mill	3894
-	346
Diffuser	232
Screw Mill	220
Tahona	103
Screw Mill Roller Mill	64
Tahona Roller Mill	39
Roller Mill Screw Mill	26
Tahona Screw Mill	10
Roller Mill Tahona	9
Diffuser Roller Mill	7
Diffuser Screw Mill	5
Roller Mill Diffuser	5
Hand Crushed	1

## Tequila Production by Distillation Method

Now let's see the distillation methods of Tequila ranked by order

Table 3: Tequila Production by Distillation Method

Distillation	Total
2x distilled	4457
	271
3x distilled	227
4x distilled	3
5x distilled	2
7x distilled	1

## Count by Still Methods

Table 4: Count by Still Methods

Still	Total
Stainless Steel Pot	1928
Copper Pot	1166
Stainless Pot w/Copper Coil	999
-	587
Column	195
Column Stainless Pot w/Copper Coil	25
Copper Pot Column	22
Copper Pot Stainless Steel Pot	14
Stainless Steel Pot Column	10
Column Copper Pot	8
Stainless Steel Pot Copper Pot	4
Copper Pot Stainless Pot w/Copper Coil	2
Filipino Style (wood/copper) Stainless Pot w/Copper Coil	1

## Count by Water Source

Table 5: Tequila Production by Water Source

Water_Source	Total
Deep well water	1159
Natural spring water	319
Natural spring water Rain water	8
Deep well water Natural spring water Rain water	4
Distilled water	4
Deep well water Rain water	3
Deep well water Natural spring water	1
Rain water Natural spring water	1

## Count by ABV

Table 6: Count by ABV

ABVorProof	Total
40% abv (80-proof)	3318
35-39% abv (70-78 proof)	626
38% abv (76 proof)	84
35% abv (70 proof)	78
35-39% abv (70-78 proof) 40% abv (80-proof)	47
50% abv (100-proof)	38
40% abv (80-proof) 35-39% abv (70-78 proof)	34
42% abv (84-proof)	32
40% abv (80-proof) 38% abv (76 proof)	30
46% abv (92-proof)	23
55% abv (110-proof)	20
43% abv (86-proof)	18
45% abv (90-proof)	12
40% abv (80-proof) 35% abv (70 proof)	8
41.5% abv (83-proof)	8



# Data Analysis and Modeling

## Ratings for overall Regions

In this analysis, we delve into the distinctions between panel and community ratings and explore the variations they exhibit. To discern these differences, I partitioned the data into two distinct sets. This segregation arose from disparities observed in how Tequila Sommeliers, industry professionals, and the general public assess and grade tequilas. Notably, I amalgamated Sommeliers and industry professionals into a unified category due to the remarkable similarity in their results, whereas the general public's ratings demonstrated considerable divergence from those of industry professionals.

Here, we commence the breakdown of the data to examine how regions are typically rated, considering panel, community, and aggregated ratings. This approach provides a comprehensive overview of the ratings attributed to each region, encompassing evaluations from diverse perspectives.

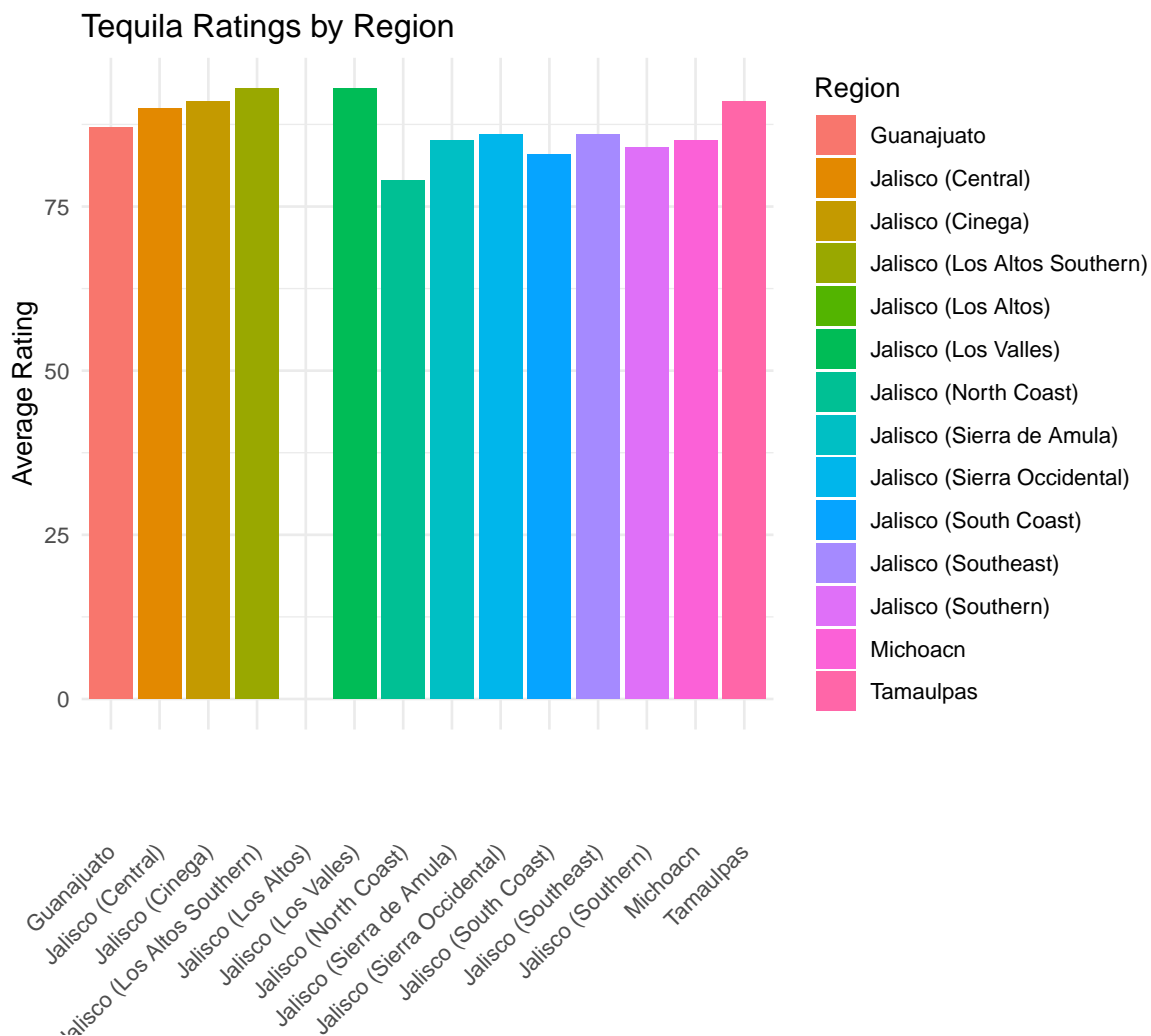


Figure 1: Tequila Ratings by Region

Table 7: Average Rating for Community Rating by Region

Region	AverageRatingCommunity
Guanajuato	80.56
Jalisco (Central)	77.30
Jalisco (Cinega)	79.87
Jalisco (Los Altos Southern)	79.41
Jalisco (Los Altos)	NaN
Jalisco (Los Valles)	80.23
Jalisco (North Coast)	76.00
Jalisco (Sierra Occidental)	85.00
Jalisco (Sierra de Amula)	89.50
Jalisco (South Coast)	79.00
Jalisco (Southeast)	78.68
Jalisco (Southern)	81.58
Michoacn	79.33
Tamaulpas	84.17

Table 8: Highest and Lowest Community Ratings by Region

Region	HighestCommunityRating	LowestCommunityRating
Guanajuato	94	45
Jalisco (Central)	100	29
Jalisco (Cinega)	98	12
Jalisco (Los Altos Southern)	100	1
Jalisco (Los Altos)	-Inf	Inf
Jalisco (Los Valles)	100	1
Jalisco (North Coast)	76	76
Jalisco (Sierra Occidental)	85	85
Jalisco (Sierra de Amula)	91	88
Jalisco (South Coast)	80	78
Jalisco (Southeast)	87	64
Jalisco (Southern)	99	61
Michoacn	88	63
Tamaulpas	94	71

## Ratings from Community and Panel raters

Here, we examine the community ratings for each region. Initially, I suspected errors in the data due to the wide spread of reported ratings. Upon careful examination, I confirmed the accuracy of the ratings. Some individuals assigned exceptionally low ratings, as certain tequilas were labeled as Mixtos rather than genuine tequila. I opted to retain these ratings, providing a realistic reflection of the diverse perspectives captured in the data.

Table 9: Average Rating for Panel Rating by Region

Region	AverageRatingPanel
Guanajuato	77.00
Jalisco (Central)	74.09
Jalisco (Cinega)	79.38
Jalisco (Los Altos Southern)	79.00
Jalisco (Los Altos)	NaN
Jalisco (Los Valles)	77.32
Jalisco (North Coast)	79.00
Jalisco (Sierra Occidental)	85.00
Jalisco (Sierra de Amula)	85.00
Jalisco (South Coast)	74.60
Jalisco (Southeast)	77.65
Jalisco (Southern)	79.62
Michoacn	75.67
Tamaulpas	77.78

Table 10: Highest and Lowest Panel Ratings by Region

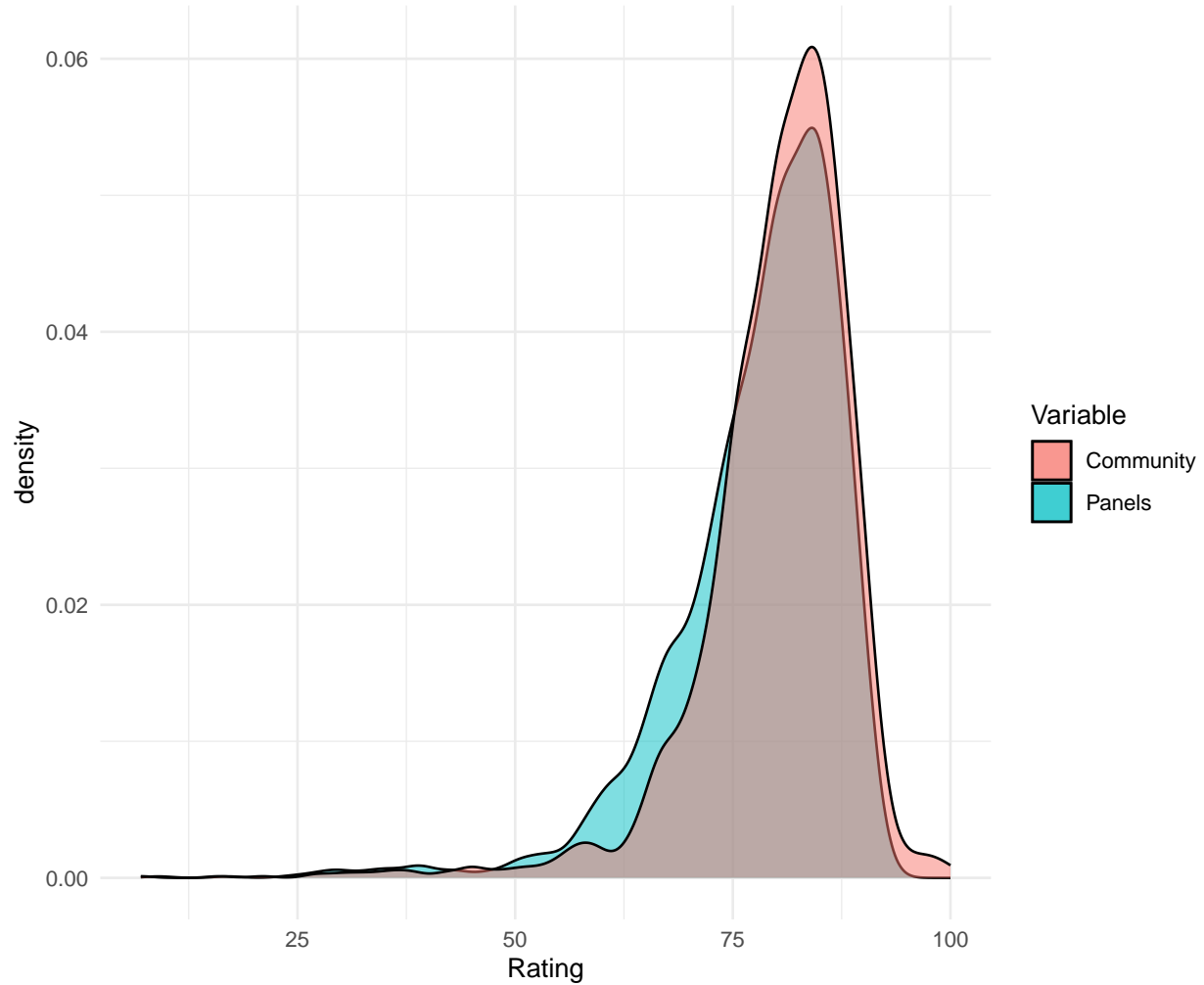
Region	HighestPanelRating	LowestPanelRating
Guanajuato	87	66
Jalisco (Central)	90	9
Jalisco (Cinega)	91	28
Jalisco (Los Altos Southern)	93	17
Jalisco (Los Altos)	-Inf	Inf
Jalisco (Los Valles)	93	10
Jalisco (North Coast)	79	79
Jalisco (Sierra Occidental)	86	84
Jalisco (Sierra de Amula)	85	85
Jalisco (South Coast)	83	63
Jalisco (Southeast)	86	63
Jalisco (Southern)	84	72
Michoacn	85	38
Tamaulpas	91	57

## Density plot ratings for the panel and community

In this analysis, we delve into the panel ratings for each region. Initially, there were concerns about data accuracy due to the extensive range of reported ratings just as reported in the community raters. However, upon meticulous examination, the ratings were verified to be accurate. It was observed that some individuals assigned notably low ratings but not as low as the community raters, attributing them to tequilas labeled as Mixtos rather than genuine tequila. Despite this variation, these ratings have been retained to offer a genuine portrayal of the diverse perspectives captured in the data.

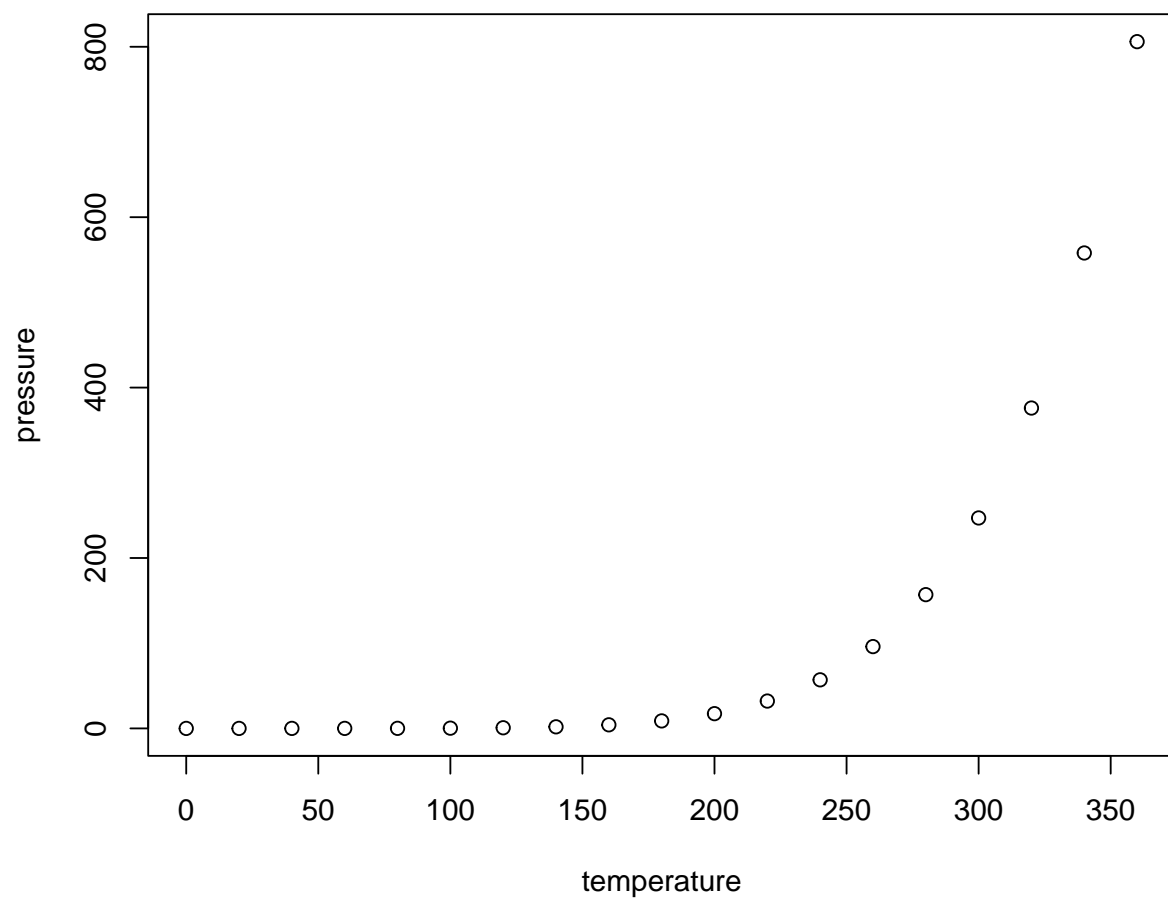
We will now explore the distinctions in ratings between the community and the professional panel using a density plot graph. This visual representation unveils the nuances in rating patterns. The community tends to assign higher ratings, showcasing a tendency towards positivity, while the professional panel exhibits a more discerning approach with a broader range of ratings, including lower scores.

### Density Plot Comparison



### Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.