Number one coolest RMD File

Matt Viana

2023-11-29

#Load packages with ground hog  
library("groundhog")

## Warning: package 'groundhog' was built under R version 4.3.2

## Attached: 'Groundhog' (Version: 3.1.2)

## Tips and troubleshooting: https://groundhogR.com

pak <- c("ggplot2", "dplyr")  
groundhog.library(pak, '2023-11-25')

## Warning: package 'ggplot2' was built under R version 4.3.2

## Warning: package 'dplyr' was built under R version 4.3.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## [36mSuccessfully attached 'ggplot2\_3.4.4'[0m

## [36mSuccessfully attached 'dplyr\_1.1.4'[0m

# Matt’s Cool RMD File

## Collatz Conjecture

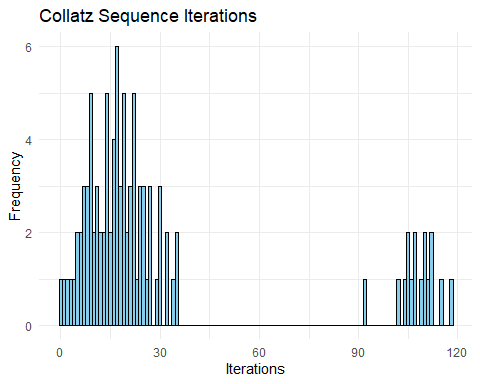
The Collatz Conjecture probes the behavior of numbers under a sequence defined by applying straightforward arithmetic operations. It ponders whether every number, when subjected to repeated divisions by 2 for even numbers and specific calculations for odd ones, eventually converges to 1. Neil sought to investigate the distribution of stopping numbers within the initial 10,000 integers, aiming to discern the frequency of iterations required for convergence.

In pursuit of this inquiry, I generated the Collatz sequence for integers ranging from 1 to 100. The resulting histogram meticulously delineates the distribution of these stopping numbers, offering a visual depiction of the frequency of iterations each integer undergoes before reaching the value 1.

The histogram shows that the distribution of the number of iterations is skewed to the left, with more numbers requiring fewer iterations to reach 1.

This is because the Collatz conjecture tends to converge relatively quickly to a small number of cycles.

# Collatz function  
collatz <- function(n, counter = -1) {  
 counter = counter + 1  
 if (n == 1) {  
 return(counter)  
 } else if (n %% 2 == 0) {  
 return(collatz((n / 2), counter))  
 } else {  
 return(collatz((3 \* n + 1), counter))  
 }  
}  
  
#Iterations  
start\_values <- 1:100  
iterations <- sapply(start\_values, collatz)  
  
  
data <- data.frame(iterations = iterations)  
  
#Information Display  
ggplot(data, aes(x = iterations)) +  
 geom\_histogram(binwidth = 1, fill = "skyblue", color = "black") +  
 labs(title = "Collatz Sequence Iterations",  
 x = "Iterations",  
 y = "Frequency") +  
 theme\_minimal()



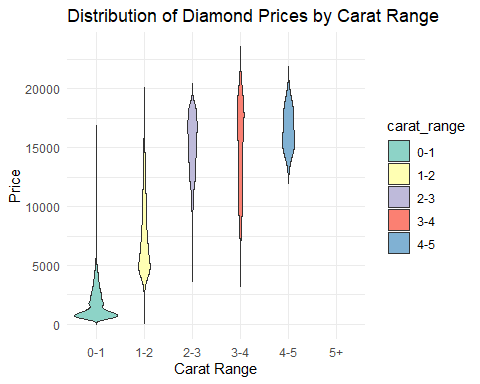
## What we can learn from Diamond Data

The graph depicts the frequency distribution of diamond carat values, revealing a skewed distribution with most diamonds weighing less than 1 carat. There is a diminishing number of diamonds with higher weights. This skewed distribution stems from the rarity leading to a rapid price increase as carat weight increases. Other factors such as clarity could influence the values, explaining the difference in price in same carat stones.

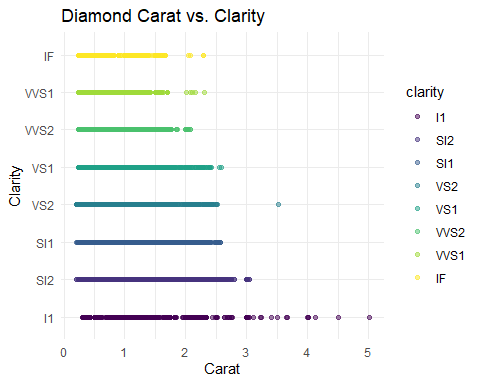
The second graph represents the distribution of carat and clarity. Together they might better explain the price of diamonds and their fluctuations.

# Start variables for ranges  
diamonds <- diamonds %>%  
 mutate(carat\_range = cut(carat, breaks = c(0, 1, 2, 3, 4, 5, Inf),  
 labels = c("0-1", "1-2", "2-3", "3-4", "4-5", "5+")))  
  
# Create violin plot  
ggplot(diamonds, aes(x = carat\_range, y = price, fill = carat\_range)) +  
 geom\_violin(trim = FALSE) +  
 scale\_fill\_brewer(palette = "Set3") +  
 labs(title = "Distribution of Diamond Prices by Carat Range",  
 x = "Carat Range",  
 y = "Price") +  
 theme\_minimal()

## Warning: Groups with fewer than two data points have been dropped.



# display carat and clarity of diamond data  
ggplot(diamonds, aes(x = carat, y = clarity, color = clarity)) +  
 geom\_point(alpha = 0.5) +  
 labs(  
 title = "Diamond Carat vs. Clarity",  
 x = "Carat",  
 y = "Clarity"  
 ) +  
 theme\_minimal()



|  |  |  |
| --- | --- | --- |
| Feature | Carat vs. Price | Carat vs. Clarity |
| Carat weights | 0 to 5.0 carats | 0.5 to 5.0 carats |
| Price range | $0 to $25,000 | - |
| Clarity grades | - | FL to I1 |
| Trends | Diamond prices increase rapidly with increasing carat weight. | Clarity grades tend to decrease with increasing carat weight. |

The first table summarizes the carat weight and price of diamonds. The carat weight ranges from 0 to 5.0 carats, and the price ranges from $0 to $25,000. With the most common range being diamonds cheaper than $5,000.

The second table summarizes the carat weight and clarity grade of diamonds. The carat weight ranges from 0 to 5.0 carats, and the clarity grade ranges from FL to I1. The largest stones are dis-proportionally less clear.

## What we learned:

Effective data visualization, exploratory data analysis, and coding in R emerged prominently. We created insightful visual representations using ggplot2 in R, learned the importance of clear and informative narratives in data explanations, and the significance of explaining data. Statistical analysis is more than just collecting and displaying data. A large portion of the job is understanding the data and deriving conclusions from them.