

SETH L.U.J and SIR M.V COLLEGE

PRAC 11 12 13 14 15

AIM:- 11 Reshaping data using pivot_longer()/pivot_wider() (R).

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for reading data, pivoting it to long format, and printing the results.
- Environment:** Lists the objects created in the session: `category_arrows` (100 obs. of 7 variables), `df` (100 obs. of 4 variables), `long_df` (300 obs. of 3 variables), `student_marks` (100 obs. of 8 variables), and `wide_df` (100 obs. of 4 variables).
- Console:** Shows the execution output, including the original wide data and the resulting long data structure.
- Files Panel:** Displays the file explorer with various project files.

```
# R Script  
library(dplyr)  
library(tidyverse)  
  
# 1. SETUP: Read student marks data and add student ID  
df <- read_csv("C:/Users/mvlu/downloads/student_marks.csv", na.strings = c("", "NA")) %>%  
  mutate(student_id = row_number()) %>%  
  select(student_id, number_courses, time_study, marks)  
  
# Print the original wide data  
print("1. Original wide data ---")  
print(head(df))  
# student_id number_courses time_study marks  
# 1 1 3 8.508 18.282  
# 2 3 4 0.098 7.732  
# 3 3 4 3.133 11.811  
# 4 4 6 7.909 11.018  
# 5 5 8 7.411 55.299  
# 6 6 9 3.211 17.822  
  
# 2. PIVOT_LONGER: Stack study metrics and marks into long format  
long_df <- df %>%  
  pivot_longer(  
    cols = c(number_courses, time_study, marks),  
    names_to = "metric",  
    values_to = "value"  
  )  
  
# Print the long format data  
print("2. Long format (pivot_longer) ---")
```

The console output shows the first six rows of the original wide data:

student_id	number_courses	time_study	marks	
1	3	8.508	18.282	
2	3	4	0.098	7.732
3	3	4	3.133	11.811
4	4	6	7.909	11.018
5	5	8	7.411	55.299
6	6	9	3.211	17.822

SETH L.U.J and SIR M.V COLLEGE

PRAC 11 12 13 14 15

The image displays two screenshots of the RStudio environment, showing R code execution and the resulting data structures.

Top Screenshot:

- Environment:** Shows a data frame with 100 observations and 4 variables: category_pivot, df, long_df, and wide_df.
- Code Editor:** Contains R code for data manipulation, including printing, pivoting, and summarizing data.
- Console:** Displays the output of the code, showing a data frame with 100 observations and 4 variables.

Bottom Screenshot:

- Environment:** Shows a data frame with 100 observations and 4 variables: category_pivot, df, long_df, and wide_df.
- Code Editor:** Contains R code for data manipulation, including printing, pivoting, and summarizing data.
- Console:** Displays the output of the code, showing a data frame with 100 observations and 4 variables.

SETH L.U.J and SIR M.V COLLEGE

PRAC 11 12 13 14 15

```

19 # 3. PIVOT WIDER: Spread data into wide format
20
21 # Print the wide data
22 print(head(wide_df, 5)) # show first 5 rows (3 metrics per student)
23
24 # 4. PIVOT WIDER: Spread data into wide format
25 wide_df <- pivot_wider(data = df,
26   names_from = 'Metric',
27   values_from = 'Value')
28
29 # Print the wide data
30 print(head(wide_df, 5))

```

Environment: History Connections Tutorial

Object	Class	Variables	Observations
category_pivot	data.frame	2	100
df	data.frame	4	100
long_df	data.frame	3	300
student_marks	data.frame	4	100
wide_df	data.frame	4	100

12 Combining datasets vertically (concatenation) using rbind() (R).

```

6 bats_clean <- data.frame(
7   Name = bats_df$species,
8   Value = bats_df$aspan_numeric
9 )
10
11 student_clean <- data.frame(
12   Name = student_df$number_courses,
13   Value = as.numeric(student_df$marks)
14 )
15
16 # Print the combined data
17 print(head(rbind(bats_clean, student_clean)))

```

Environment: History Connections Tutorial

Object	Class	Variables	Observations
bats_df	data.frame	11	30000
bats_information	data.frame	11	30000
category_pivot	data.frame	2	100
combined_data	data.frame	2	30100
df	data.frame	4	100
long_df	data.frame	3	300
student_clean	data.frame	2	100
student_marks	data.frame	4	100
wide_df	data.frame	4	100

AIM:- 13 Identifying and handling duplicates using distinct() (R).

OUTPUT:-

The screenshot shows the RStudio interface with the following content:

Console:

```
R> R632 -->
> bats_df <- read.csv("C:/Users/mvliet/Downloads/bats_information.csv")
>
> bats_clean <- data.frame(
+   Name = bats_df$Species,
+   Value = as.numeric(sub("-", "", bats_df$Average.Length))
+ )
Warning message:
In data.frame(Name = bats_df$Species, Value = as.numeric(sub("-", "",
NA: introduced by coercion
> print("---- Original Data ----")
[1] "---- Original Data ----"
> print(bats_clean)
```

Environment:

Object	Class	Size
Global Environment		124 MB
combined_data	data.frame	10100 obs. of 2 variables
df	data.frame	100 obs. of 4 variables
duplicates_report	data.frame	18 obs. of 1 variable
long_df	data.frame	300 obs. of 3 variables
student_clean	data.frame	100 obs. of 2 variables
student_df	data.frame	100 obs. of 1 variable
Student_Marks	data.frame	100 obs. of 1 variable

The screenshot shows the RStudio interface with the following content:

Console:

```
R 4.2.3 --
> read_csv("bats.csv")
# A tibble: 18 x 3
  name      name_value [NA] value      n
<chr>      <chr>      <dbl> <dbl>
1 Yangochiroptera (Microbat); Example: Big Brown Bat NA 1672
2 Yangochiroptera (Microbat); Example: Brandt's Bat NA 1687
3 Yangochiroptera (Microbat); Example: Bumblebee Bat NA 1690
4 Yangochiroptera (Microbat); Example: Egyptian Fruit Bat NA 1709
5 Yangochiroptera (Microbat); Example: Fringed Myotis NA 1749
6 Yangochiroptera (Microbat); Example: Horseshoe Bat NA 1636
7 Yangochiroptera (Microbat); Example: Little Brown Bat NA 1644
8 Yangochiroptera (Microbat); Example: Spectral Bat NA 1392
9 Yangochiroptera (Microbat); Example: Vampire Bat NA 1696
10 Vesperugo (Megabat); Example: Big Brown Bat NA 1693
11 Vesperugo (Megabat); Example: Brandt's Bat NA 1723
12 Vesperugo (Megabat); Example: Bumblebee Bat NA 1617
13 Vesperugo (Megabat); Example: Egyptian Fruit Bat NA 1657
14 Vesperugo (Megabat); Example: Fringed Myotis NA 1732
15 Vesperugo (Megabat); Example: Horseshoe Bat NA 1653
```

Environment:

Object	Class	Size	Variables
combined_data	tbl_df, tbl	10100 obs.	of 2 variables
df	tbl_df, tbl	100 obs.	of 4 variables
duplicates_report	tbl_df, tbl	18 obs.	of 3 variables
long_df	tbl_df, tbl	300 obs.	of 3 variables
student_clean	tbl_df, tbl	100 obs.	of 2 variables
student_df	tbl_df, tbl	100 obs.	of 3 variables
Student_Marks	tbl_df, tbl	100 obs.	of 3 variables

SETH L.U.J and SIR M.V COLLEGE

PRAC 11 12 13 14 15

The first screenshot shows the RStudio interface with the following code in the console:

```
R - R4.2.3 --
> library(lubridate)
> library(dplyr)
> bats_clean %>%
  distinct()
# A tibble: 18 x 2
#   Name                                     Value
#   <chr>                                <dbl>
1 Yangochiroptera (Microbat); Example: Big Brown Bat      NA
2 Yangochiroptera (Microbat); Example: Brandt's Bat       NA
3 Yangochiroptera (Microbat); Example: Bumblebee Bat      NA
4 Yangochiroptera (Microbat); Example: Egyptian Fruit Bat NA
5 Yangochiroptera (Microbat); Example: Fringed Myotis     NA
6 Yangochiroptera (Microbat); Example: Horseshoe Bat      NA
7 Yangochiroptera (Microbat); Example: Little Brown Bat   NA
8 Yangochiroptera (Microbat); Example: Spectral Bat       NA
9 Yangochiroptera (Microbat); Example: Vampire Bat       NA
10 Vinyterochiroptera (Megabat); Example: Big Brown Bat    NA
11 Vinyterochiroptera (Megabat); Example: Brandt's Bat     NA
12 Vinyterochiroptera (Megabat); Example: Bumblebee Bat   NA
13 Vinyterochiroptera (Megabat); Example: Egyptian Fruit Bat NA
14 Vinyterochiroptera (Megabat); Example: Fringed Myotis  NA
15 Vinyterochiroptera (Megabat); Example: Horseshoe Bat   NA
16 Vinyterochiroptera (Megabat); Example: Little Brown Bat NA
17 Vinyterochiroptera (Megabat); Example: Spectral Bat    NA
18 Vinyterochiroptera (Megabat); Example: Vampire Bat    NA
```

The second screenshot shows the same RStudio interface with the following code in the console:

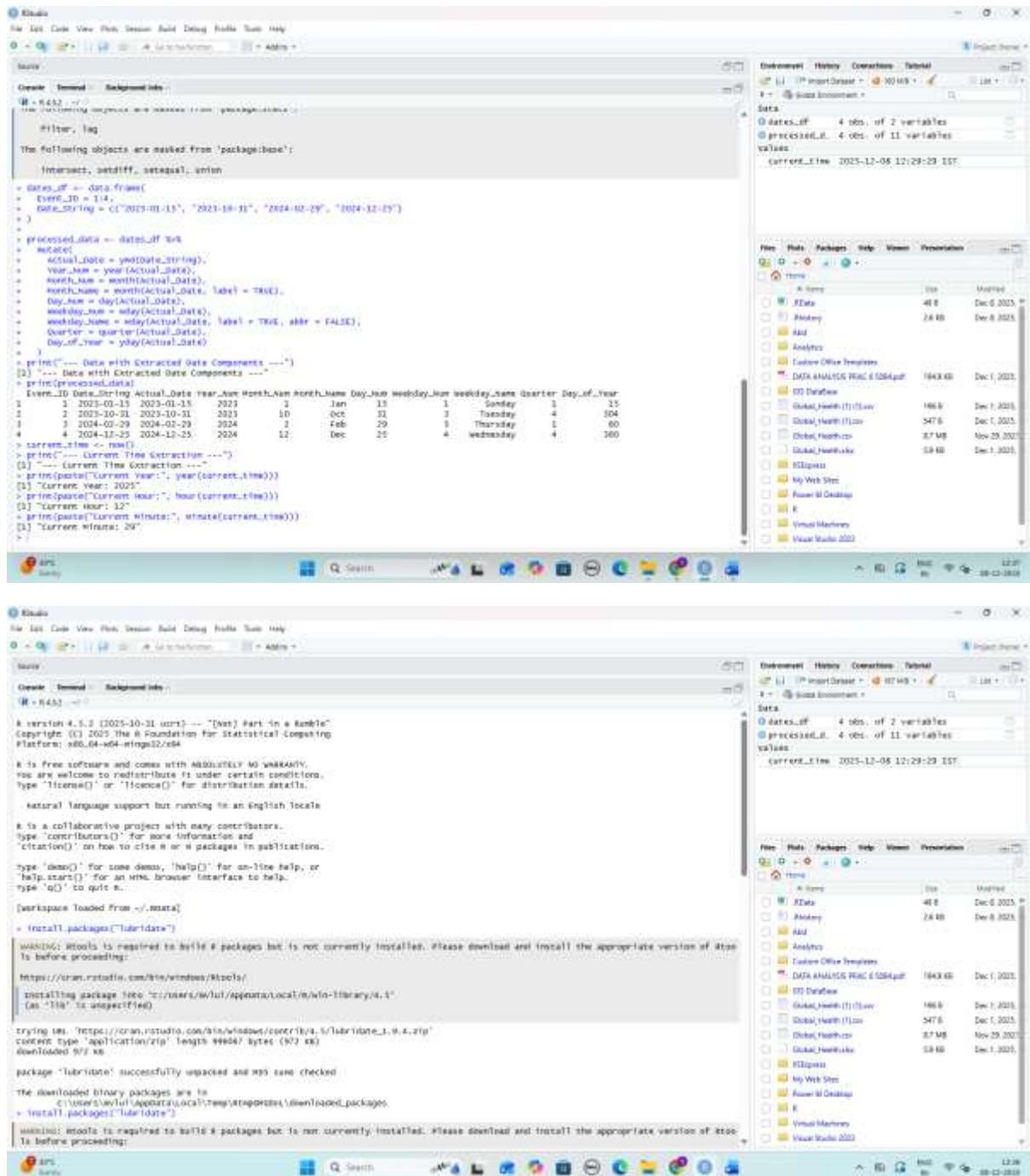
```
R - R4.2.3 --
> bats_clean %>%
  distinct()
# A tibble: 18 x 2
#   Name                                     Value
#   <chr>                                <dbl>
1 Yangochiroptera (Microbat); Example: Big Brown Bat      NA
2 Yangochiroptera (Microbat); Example: Brandt's Bat       NA
3 Yangochiroptera (Microbat); Example: Bumblebee Bat      NA
4 Yangochiroptera (Microbat); Example: Egyptian Fruit Bat NA
5 Yangochiroptera (Microbat); Example: Fringed Myotis     NA
6 Yangochiroptera (Microbat); Example: Horseshoe Bat      NA
7 Yangochiroptera (Microbat); Example: Little Brown Bat   NA
8 Yangochiroptera (Microbat); Example: Spectral Bat       NA
9 Yangochiroptera (Microbat); Example: Vampire Bat       NA
10 Vinyterochiroptera (Megabat); Example: Big Brown Bat    NA
11 Vinyterochiroptera (Megabat); Example: Brandt's Bat     NA
12 Vinyterochiroptera (Megabat); Example: Bumblebee Bat   NA
13 Vinyterochiroptera (Megabat); Example: Egyptian Fruit Bat NA
14 Vinyterochiroptera (Megabat); Example: Fringed Myotis  NA
15 Vinyterochiroptera (Megabat); Example: Horseshoe Bat   NA
16 Vinyterochiroptera (Megabat); Example: Little Brown Bat NA
17 Vinyterochiroptera (Megabat); Example: Spectral Bat    NA
18 Vinyterochiroptera (Megabat); Example: Vampire Bat    NA
```

14 Extracting date components using lubridate:: functions (R).

OUTPUT:-

SETH L.U.J and SIR M.V COLLEGE

PRAC 11 12 13 14 15



15 Generating basic summaries using str() or summary() (R).

OUTPUT:-

The screenshot displays a Jupyter Notebook with the following content:

```

# Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load Data
df = pd.read_csv('data.csv')

# Display the first few rows
df.head()

# Summary Statistics
df.describe()

# Data Types
df.dtypes

# Missing Values
df.isnull().sum()

# Correlation Matrix
df.corr()

# Scatter Plot
sns.pairplot(df)

# Box Plot
sns.boxplot(df['Price'])

# Histogram
sns.histplot(df['Price'])

# Facet Grid
sns.relplot(data=df, x='Category', y='Price', hue='Brand')

# Linear Regression
model = LinearRegression()
model.fit(df[['Length', 'Width', 'Height'], df['Price']))

# Predict Prices
df['Predicted Price'] = model.predict(df[['Length', 'Width', 'Height']])

# Residuals
df['Residuals'] = df['Price'] - df['Predicted Price']

# Summary of Residuals
df['Residuals'].describe()

# Plot Residuals
sns.residplot(df['Predicted Price'], df['Residuals'])

# Conclusion
print("Data loaded successfully.")

```

The output shows the first five rows of the dataset, summary statistics, data types, missing values, correlation matrix, and various plots including a pair plot, box plot, histogram, and a faceted linear regression plot.

[illegible]