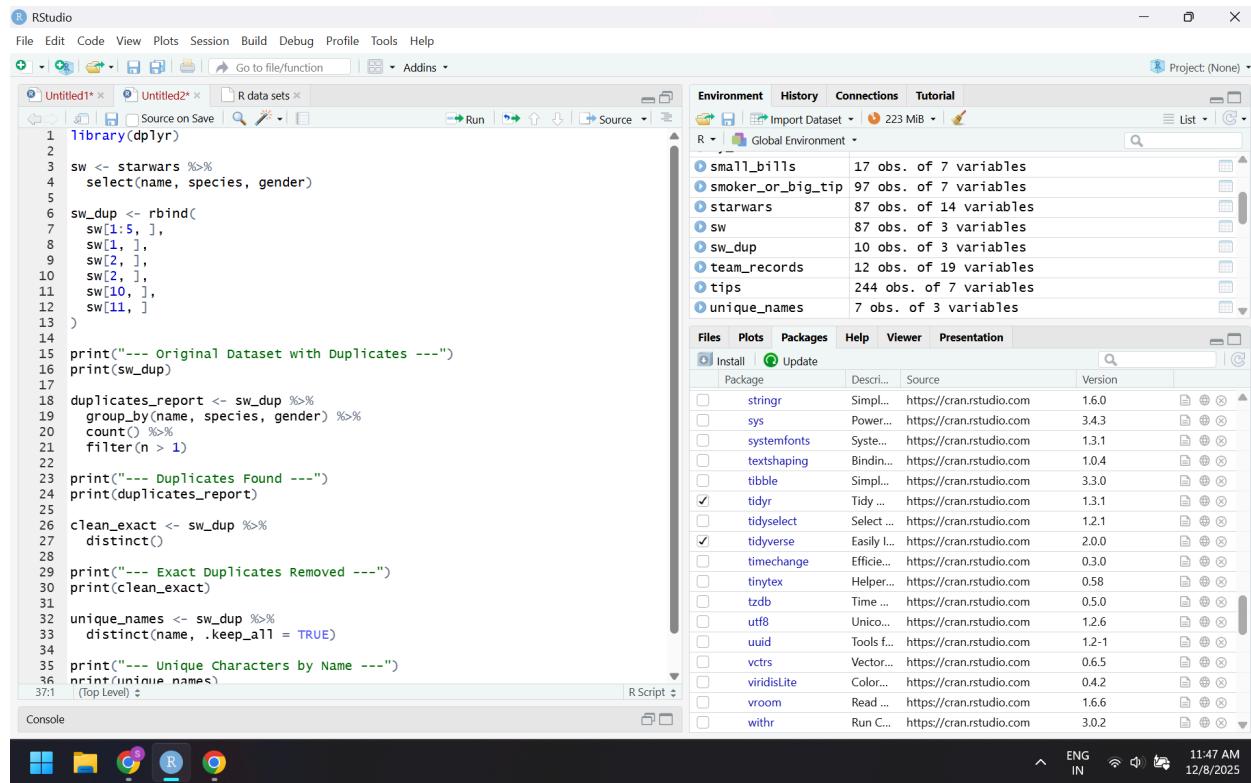


SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE

SUBJECT - Data Analysis with SAS / SPSS / R

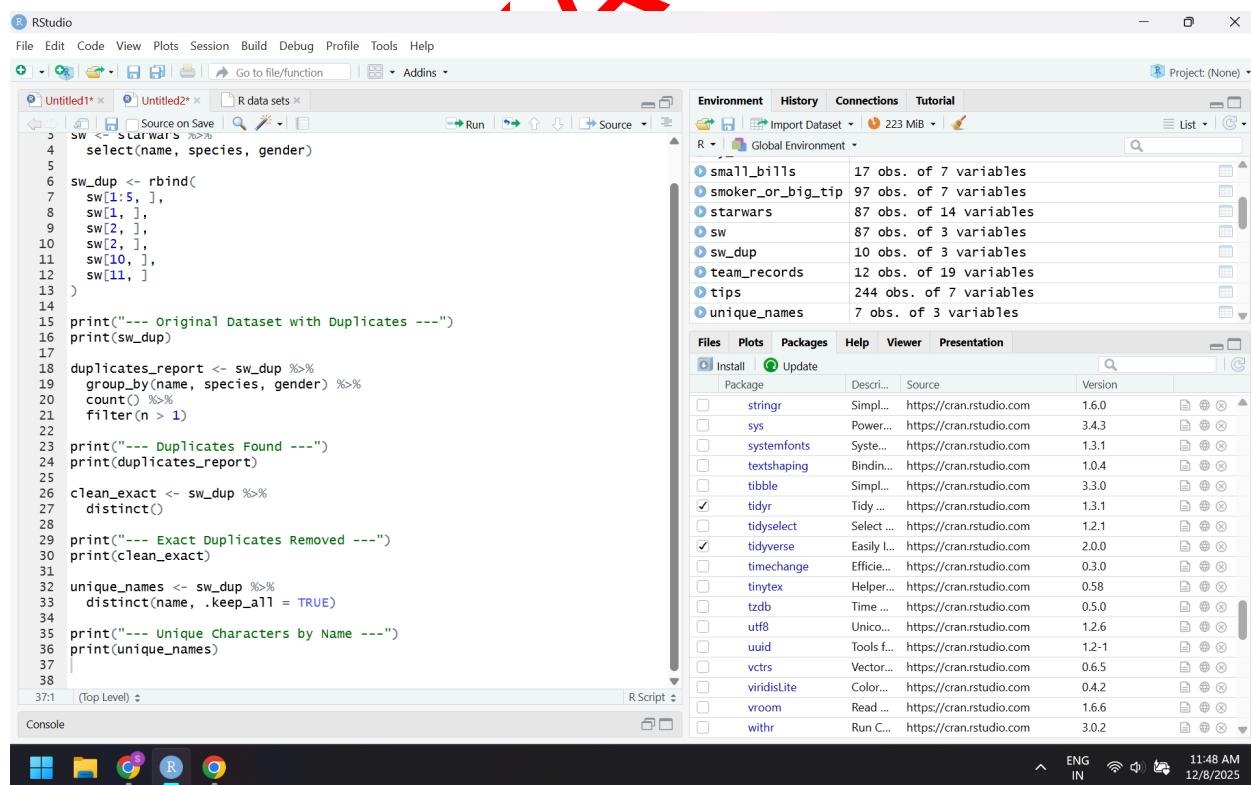
Aim :- Identifying and handling duplicates using distinct() (R studio).

Input :



```

library(dplyr)
sw <- starwars %>%
  select(name, species, gender)
sw_dup <- rbind(
  sw[1:5, ],
  sw[1, ],
  sw[2, ],
  sw[2, ],
  sw[10, ],
  sw[11, ]
)
print("---- Original Dataset with Duplicates ---")
print(sw_dup)
duplicates_report <- sw_dup %>%
  group_by(name, species, gender) %>%
  count() %>%
  filter(n > 1)
print("---- Duplicates Found ---")
print(duplicates_report)
clean_exact <- sw_dup %>%
  distinct()
print("---- Exact Duplicates Removed ---")
print(clean_exact)
unique_names <- sw_dup %>%
  distinct(name, .keep_all = TRUE)
print("---- Unique Characters by Name ---")
print(unique_names)
  
```



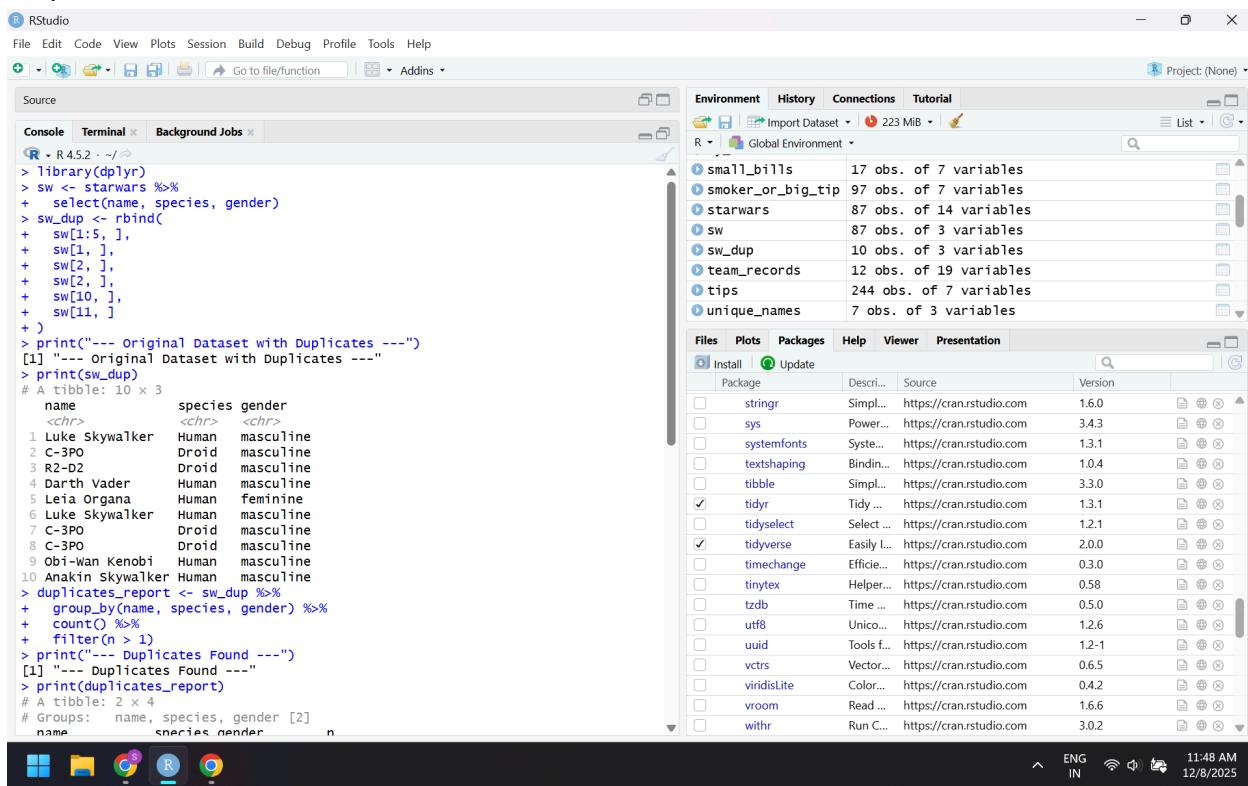
```

library(dplyr)
sw <- starwars %>%
  select(name, species, gender)
sw_dup <- rbind(
  sw[1:5, ],
  sw[1, ],
  sw[2, ],
  sw[2, ],
  sw[10, ],
  sw[11, ]
)
print("---- Original Dataset with Duplicates ---")
print(sw_dup)
duplicates_report <- sw_dup %>%
  group_by(name, species, gender) %>%
  count() %>%
  filter(n > 1)
print("---- Duplicates Found ---")
print(duplicates_report)
clean_exact <- sw_dup %>%
  distinct()
print("---- Exact Duplicates Removed ---")
print(clean_exact)
unique_names <- sw_dup %>%
  distinct(name, .keep_all = TRUE)
print("---- Unique Characters by Name ---")
print(unique_names)
  
```

Name - Mithil Kadam
 Roll No - S083

SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE
SUBJECT - Data Analysis with SAS / SPSS / R

Output :



RStudio interface showing R code execution and package management.

```

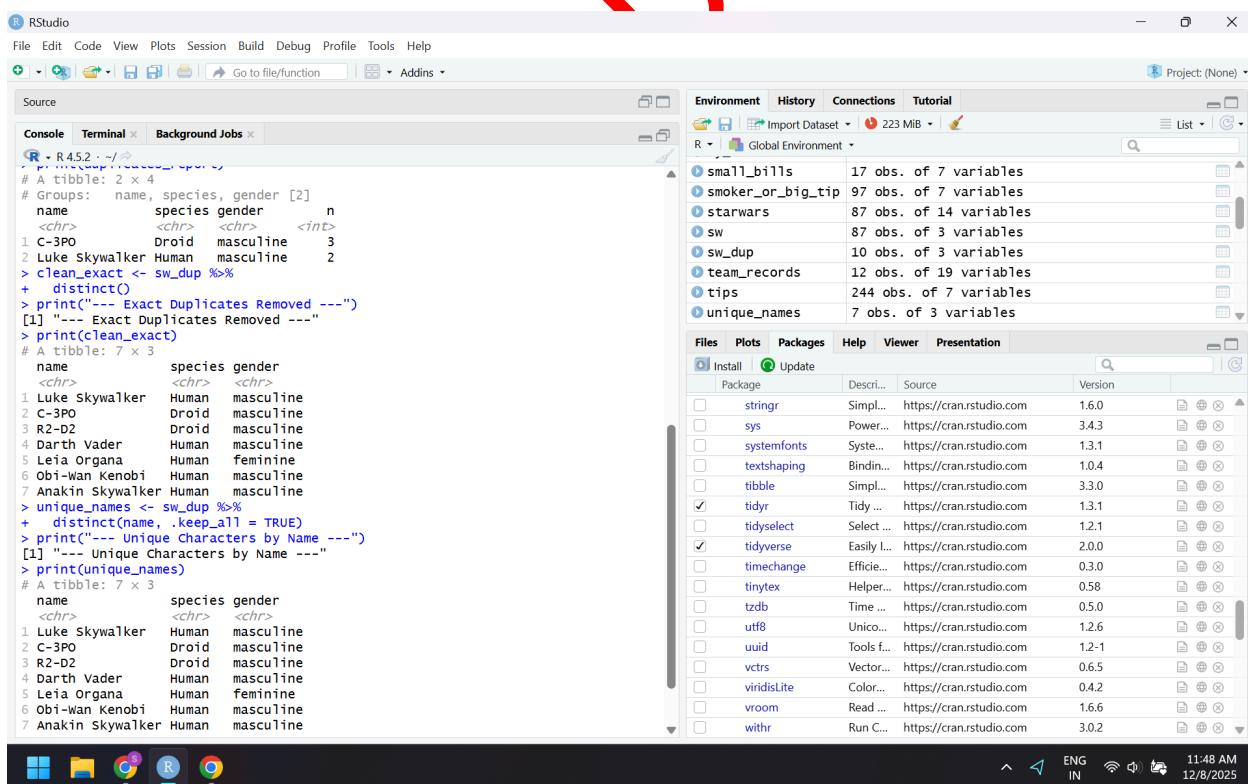
R - R 4.5.2 · ~/Documents
> library(dplyr)
> sw <- starwars %>%
+   select(name, species, gender)
> sw_dup <- rbind(
+   sw[1:5, ],
+   sw[1, ],
+   sw[2, ],
+   sw[2, ],
+   sw[10, ],
+   sw[11, ]
+ )
> print("---- Original Dataset with Duplicates ----")
[1] "---- Original Dataset with Duplicates ----"
> print(sw_dup)
# A tibble: 10 × 3
  name      species gender
  <chr>    <chr>   <chr>
1 Luke Skywalker Human   masculine
2 C-3PO     Droid    masculine
3 R2-D2    Droid    masculine
4 Darth Vader Human   masculine
5 Leia Organa Human   feminine
6 Luke Skywalker Human   masculine
7 C-3PO     Droid    masculine
8 C-3PO     Droid    masculine
9 Obi-wan Kenobi Human   masculine
10 Anakin Skywalker Human  masculine
> duplicates_report <- sw_dup %>%
+   group_by(name, species, gender) %>%
+   count() %>%
+   filter(n > 1)
> print("---- Duplicates Found ----")
[1] "---- Duplicates Found ----"
> print(duplicates_report)
# A tibble: 2 × 4
  name      species gender n
  <chr>    <chr>   <chr> <int>
1 C-3PO     Droid    masculine  3
2 Luke Skywalker Human   masculine  2
  
```

Environment pane shows available datasets:

- small_bills
- smoker_or_big_tip
- starwars
- sw
- sw_dup
- team_records
- tips
- unique_names

Packages pane shows installed packages:

Package	Description	Source	Version
stringr	Simpl...	https://cran.rstudio.com	1.6.0
sys	Power...	https://cran.rstudio.com	3.4.3
systemfonts	Syste...	https://cran.rstudio.com	1.3.1
textshaping	Bindin...	https://cran.rstudio.com	1.0.4
tibble	Simpl...	https://cran.rstudio.com	3.3.0
tidy	Tidy ...	https://cran.rstudio.com	1.3.1
tidyselect	Select ...	https://cran.rstudio.com	1.2.1
tidyverse	Easily ...	https://cran.rstudio.com	2.0.0
timechange	Efficie...	https://cran.rstudio.com	0.3.0
tinytex	Helper...	https://cran.rstudio.com	0.58
tzdb	Time ...	https://cran.rstudio.com	0.5.0
utf8	Unico...	https://cran.rstudio.com	1.2.6
uuid	Tools f...	https://cran.rstudio.com	1.2-1
vctrs	Vector...	https://cran.rstudio.com	0.6.5
viridisLite	Color...	https://cran.rstudio.com	0.4.2
vroom	Read ...	https://cran.rstudio.com	1.6.6
withr	Run C...	https://cran.rstudio.com	3.0.2



RStudio interface showing R code execution and package management.

```

R - R 4.5.2 · ~/Documents
# A tibble: 2 × 4
# Groups: name, species, gender [2]
  name      species gender n
  <chr>    <chr>   <chr> <int>
1 C-3PO     Droid    masculine  3
2 Luke Skywalker Human   masculine  2
> clean_exact <- sw_dup %>%
+   distinct()
> print("---- Exact Duplicates Removed ----")
[1] "---- Exact Duplicates Removed ----"
> print(clean_exact)
# A tibble: 7 × 3
  name      species gender
  <chr>    <chr>   <chr>
1 Luke Skywalker Human   masculine
2 C-3PO     Droid    masculine
3 R2-D2    Droid    masculine
4 Darth Vader Human   masculine
5 Leia Organa Human   feminine
6 Obi-wan Kenobi Human   masculine
7 Anakin Skywalker Human  masculine
> unique_names <- sw_dup %>%
+   distinct(name, .keep_all = TRUE)
> print("---- Unique Characters by Name ----")
[1] "---- Unique Characters by Name ----"
> print(unique_names)
# A tibble: 7 × 3
  name      species gender
  <chr>    <chr>   <chr>
1 Luke Skywalker Human   masculine
2 C-3PO     Droid    masculine
3 R2-D2    Droid    masculine
4 Darth Vader Human   masculine
5 Leia Organa Human   feminine
6 Obi-wan Kenobi Human   masculine
7 Anakin Skywalker Human  masculine
  
```

Environment pane shows available datasets:

- small_bills
- smoker_or_big_tip
- starwars
- sw
- sw_dup
- team_records
- tips
- unique_names

Packages pane shows installed packages:

Package	Description	Source	Version
stringr	Simpl...	https://cran.rstudio.com	1.6.0
sys	Power...	https://cran.rstudio.com	3.4.3
systemfonts	Syste...	https://cran.rstudio.com	1.3.1
textshaping	Bindin...	https://cran.rstudio.com	1.0.4
tibble	Simpl...	https://cran.rstudio.com	3.3.0
tidy	Tidy ...	https://cran.rstudio.com	1.3.1
tidyselect	Select ...	https://cran.rstudio.com	1.2.1
tidyverse	Easily ...	https://cran.rstudio.com	2.0.0
timechange	Efficie...	https://cran.rstudio.com	0.3.0
tinytex	Helper...	https://cran.rstudio.com	0.58
tzdb	Time ...	https://cran.rstudio.com	0.5.0
utf8	Unico...	https://cran.rstudio.com	1.2.6
uuid	Tools f...	https://cran.rstudio.com	1.2-1
vctrs	Vector...	https://cran.rstudio.com	0.6.5
viridisLite	Color...	https://cran.rstudio.com	0.4.2
vroom	Read ...	https://cran.rstudio.com	1.6.6
withr	Run C...	https://cran.rstudio.com	3.0.2