R tidyverse to SAS – Practical Equivalence Guide

This document maps commonly used R tidyverse verbs and functions to their closest SAS equivalents. It is designed for day-to-day data analysis translation between R and SAS.

1. Core Data Manipulation (dplyr)

R Verb	Purpose	SAS Equivalent
select()	Pick specific columns	KEEP= / DROP=
filter()	Filter rows by condition	WHERE / IF
arrange()	Sort rows	PROC SORT
mutate()	Create or modify columns	DATA step assignments
summarise()	Aggregate data	PROC SUMMARY / PROC SQL
group_by()	Define groups	CLASS / BY / GROUP BY
rename()	Rename columns	RENAME=
distinct()	Remove duplicates	PROC SORT NODUPKEY
count()	Count rows by group	PROC FREQ / PROC SUMMARY
slice()	Select rows by position	_N_, FIRST./LAST.
pull()	Extract column as vector	CALL SYMPUT / macro var

2. Combining & Joining Data

R Verb	Purpose	SAS Equivalent
left_join()	Keep all rows from left	PROC SQL LEFT JOIN
right_join()	Keep all rows from right	PROC SQL RIGHT JOIN
inner_join()	Matching rows only	PROC SQL INNER JOIN
full_join()	All rows from both	PROC SQL FULL JOIN
bind_rows()	Stack datasets	SET statement
bind_cols()	Combine columns	MERGE (no BY)

3. Reshaping Data (tidyr)

R Verb	Purpose	SAS Equivalent
pivot_longer()	Wide to long	PROC TRANSPOSE
pivot_wider()	Long to wide	PROC TRANSPOSE

separate()	Split column	SCAN()
unite()	Combine columns	CATX()
drop_na()	Remove missing rows	IF NOT MISSING()
replace_na()	Replace missing	IF MISSING()

4. Data Cleaning & Helpers

R Function	Purpose	SAS Equivalent
across()	Apply across columns	Arrays
if_else()	Conditional logic	IF THEN ELSE
case_when()	Multiple conditions	SELECT WHEN
coalesce()	First non-missing	COALESCE()
rowwise()	Row-wise operations	Default SAS behavior

5. Date & Time Functions

R Function	Purpose	SAS Equivalent
ymd()/mdy()/dmy()	Parse dates	INPUT(..., date informat)
year()	Extract year	YEAR()
month()	Extract month	MONTH()
day()	Extract day	DAY()
wday()	Weekday	WEEKDAY()
qtr()	Quarter	QTR()
floor_date()	Round down	INTNX()
ceiling_date()	Round up	INTNX()
difftime()	Date difference	INTCK()
today()	Current date	TODAY()

6. String Functions

R Function	Purpose	SAS Equivalent
str_detect()	Pattern match	PRXMATCH / INDEX
str_replace()	Replace first match	PRXCHANGE
str_replace_all()	Replace all matches	TRANWRD

str_extract()	Extract pattern	PRXCHANGE
str_sub()	Substring	SUBSTR
str_length()	String length	LENGTH
str_to_upper()	Uppercase	UPCASE
str_to_lower()	Lowercase	LOWCASE
str_trim()	Trim spaces	STRIP
paste()	Join with separator	CATX
paste0()	Join without separator	CATS
sprintf()	Formatted string	PUT

7. Numeric Functions

R Function	Purpose	SAS Equivalent
mean()	Average	MEAN
median()	Median	MEDIAN
sum()	Total	SUM
min()/max()	Min / Max	MIN / MAX
round()	Round	ROUND
ceiling()	Round up	CEIL
floor()	Round down	FLOOR
sd()	Std deviation	STD
var()	Variance	VAR
n()	Row count	_N_ / COUNT(*)
n_distinct()	Distinct count	COUNT(DISTINCT)
rank()	Rank values	PROC RANK

8. Functions Used Everywhere

R Function	Purpose	SAS Equivalent
is.na()	Check missing	MISSING()
if_else()	Conditional logic	IF THEN ELSE
case_when()	Multi-condition logic	SELECT WHEN
coalesce()	First non-missing	COALESCE()

na.rm=TRUE	Ignore missing	Default SAS behavior

Key Takeaway

In practice, most tidyverse pipelines translate into a combination of DATA steps, PROC SQL, and PROC SUMMARY in SAS. Understanding this mapping allows seamless transition between R and SAS in real-world analytics.