Python ↔ R Data Wrangling Cheat Sheet (Side■by■Side Layout)

Select Columns

R	Python
`df %>% select(name, score) df %>% select(starts_with("s"))`	`df[["name", "score"]] df.loc[:, df.columns.str.startswith("s")]`

Filter Rows

R	Python
`df %>% filter(age > 24, grp == "g1")`	`df[(df["age"] > 24) & (df["grp"] == "g1")]`

Mutate / Create Columns

R	Python
`df %>% mutate( passed = score > 90, score_z = (score - mean(score)) / sd(score) )`	`df.assign( passed = df["score"] > 90, score_z = (df["score"] - df["score"].mean()) / df["score"].std() )`

Group and Summarize

R	Python
`df %>% group_by(grp) %>% summarise( n = n(), mean_score = mean(score), max_age = max(age), .groups = "drop" )`	`df.groupby("grp").agg( n=("grp", "size"), mean_score=("score", "mean"), max_age=("age", "max") ).reset_index()`

Join

R	Python
`lookup <- tibble(grp=c("g1","g2"), label=c("Group 1","Group 2")) df %>% left_join(lookup, by="grp")`	`lookup = pd.DataFrame({ "grp":["g1","g2"], "label":["Group 1","Group 2"] }) df.merge(lookup, on="grp", how="left")`