Cheat Sheet: Common Tasks in SAS, R, and Python

Task	SAS	R (dplyr)	Python (pandas)
Read Data	`proc import datafile="path" out=ds;`	`read.csv("path")`	`pd.read_csv("path")`
Filter Rows	`where var = 'value';`	`filter(var == 'value')`	`df[df['var'] == 'value']`
Select Columns	`keep var1 var2;`	`select(var1, var2)`	`df[['var1', 'var2']]`
Create New Column	`data ds; set ds; new_var = var1 + var2;`	`mutate(new_var = var1 + var2)`	`df['new_var'] = df['var1'] + df['var2']`
Group By	`proc means; class var;`	`group_by(var) %>% summarize()`	`df.groupby('var').agg()`
Summarize Data	`proc summary; var var1;`	`summarize(mean_var = mean(var1))`	`df['var1'].mean()`
Sort Data	`proc sort data=ds; by var;`	`arrange(var)`	`df.sort_values('var')`
Join/Merge Data	`merge ds1 ds2; by var;`	`left_join(ds1, ds2, by = "var")`	`pd.merge(df1, df2, on='var', how='left')`
Rename Column	`rename var1=new_var1;`	`rename(new_var1 = var1)`	`df.rename(columns={'var1':'new_var1'})`
Remove Duplicates	`proc sort nodupkey;`	`distinct()`	`df.drop_duplicates()`
Basic Plot	`proc sgplot; scatter x=var1 y=var2;`	`ggplot(data, aes(x = var1, y = var2)) + geom_point()`	`plt.scatter(df['var1'], df['var2']`