# Exercise

Use the **AmesHousing** data set to complete this exercise.

Using Descriptive Procedures and ODS

a. Navigate to the SAS Help documentation and view the TABLES statement options for the FREQ procedure. Which option enables you to create a frequency plot? Use PROC FREQ to create one-way frequency tables for the variables **central_air** and **house_style** along with frequency plots. What percentage of homes in this sample have central air? What percent are only one story?

Central_ Air	Frequency	Percent	Cumulative Frequency	Cumulative Percent
N	42	14.00	42	14.00
Y	258	86.00	300	100.00
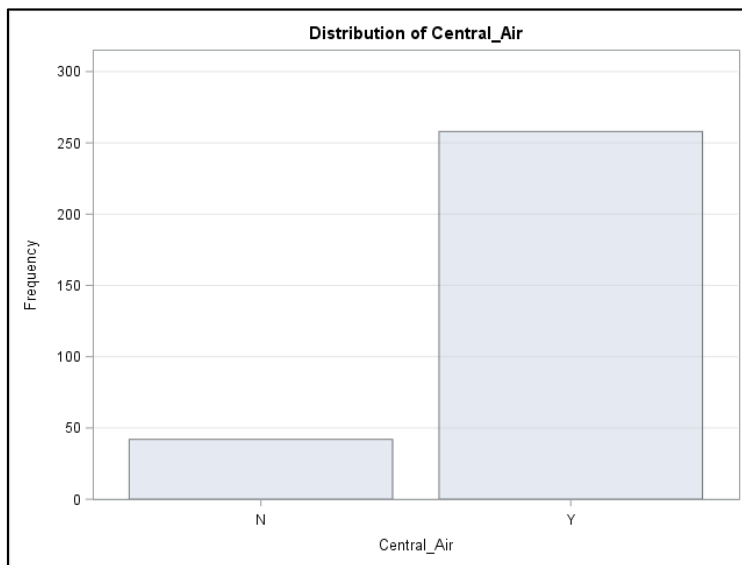
House_ Style	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.5Fin	28	9.33	28	9.33
1.5Unf	4	1.33	32	10.67
1Story	194	64.67	226	75.33
2.5Unf	2	0.67	228	76.00
2Story	38	12.67	266	88.67
SFoyer	13	4.33	279	93.00
SLvl	21	7.00	300	100.00

Distribution of House_Style
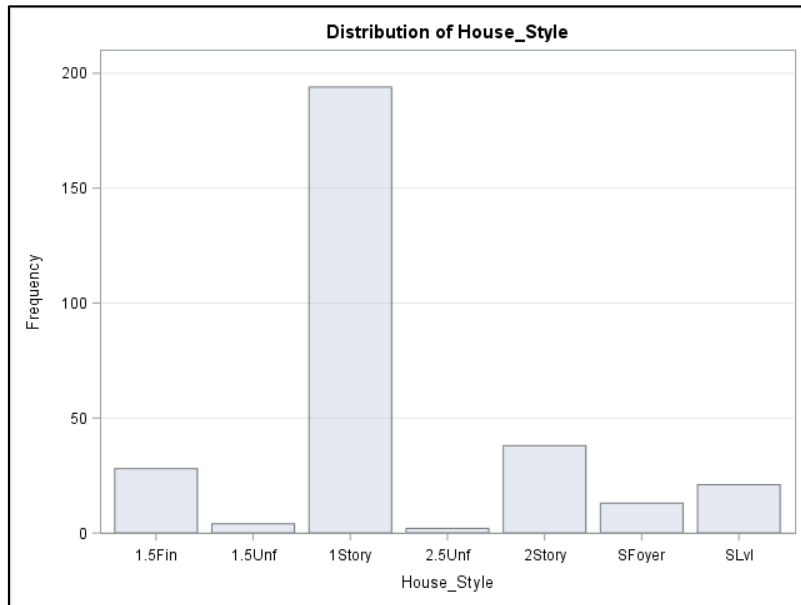
b. The default PROC CORR output gives a table of simple statistics and correlation coefficients. Use ODS SELECT to print only the correlation coefficients for the variables **saleprice**, **garage_area**, **basement_area**, and **gr_liv_area**. (Hint: It might be easiest to use the ODS TRACE statement to learn the table name instead of going to the documentation page.) Is there a statistically significant correlation between **saleprice** and each of the other variables?

	Pearson Correlation Coefficients, N = 300 Prob > |r| under H0: Rho=0			
	Sale Price	Garage_ Area	Basement_ Area	Gr_Liv_ Area
SalePrice	1.00000	0.57892 <.0001	0.68956 <.0001	0.65046 <.0001
Garage_Area	0.57892 <.0001	1.00000	0.35630 <.0001	0.33283 <.0001
Basement_Area	0.68956 <.0001	0.35630 <.0001	1.00000	0.43985 <.0001
Gr_Liv_Area	0.65046 <.0001	0.33283 <.0001	0.43985 <.0001	1.00000

c. Use PROC MEANS to print the 10th percentile, median, and 90th percentile for the variables
saleprice and **gr_liv_area**. In addition, use the CLASS statement to separate the summary
statistics by the **yr_sold** variable. Finally, save the output using ODS OUTPUT and name the
table **summary_table**. Print the table to ensure it is saved. Which year had the highest median
sale price?

Obs	Yr_Sold	NObs	VName_ SalePrice	SalePrice_ P10	SalePrice_ Median	SalePrice_ P90
1	2006	55	SalePrice	93500	131000	169000
2	2007	72	SalePrice	96500	128500	180500
3	2008	62	SalePrice	87000	136250	181900
4	2009	73	SalePrice	91300	144000	192000
5	2010	38	SalePrice	100000	148875	192000

Obs	VName_Gr_ Liv_Area	Gr_Liv_Area_ P10	Gr_Liv_Area_ Median	Gr_Liv_Area_ P90
1	Gr_Liv_Area	864	1092	1368
2	Gr_Liv_Area	864	1076	1435
3	Gr_Liv_Area	864	1185	1430
4	Gr_Liv_Area	800	1210	1456
5	Gr_Liv_Area	848	1148.5	1395

d. Use PROC UNIVARIATE to analyze the **gr_liv_area** variable and create both a histogram and a
Q-Q plot. For the histogram, overlay a normal and density kernel estimate. Use the OUTPUT
statement to create a new data table of percentiles called **gr_percs**. Instead of providing the
PCTLPTS= option a list, use the following syntax: **PCTLPTS= 40 to 60 by 2**. Let the
prefixes for the saved percentiles be **gr_**. Print the table to ensure that it is saved.

Obs	gr_40	gr_42	gr_44	gr_46	gr_48	gr_50	gr_52	gr_54	gr_56	gr_58	gr_60
1	1063.5	1075.5	1087	1092	1109.5	1135	1151	1169.5	1191	1206	1218