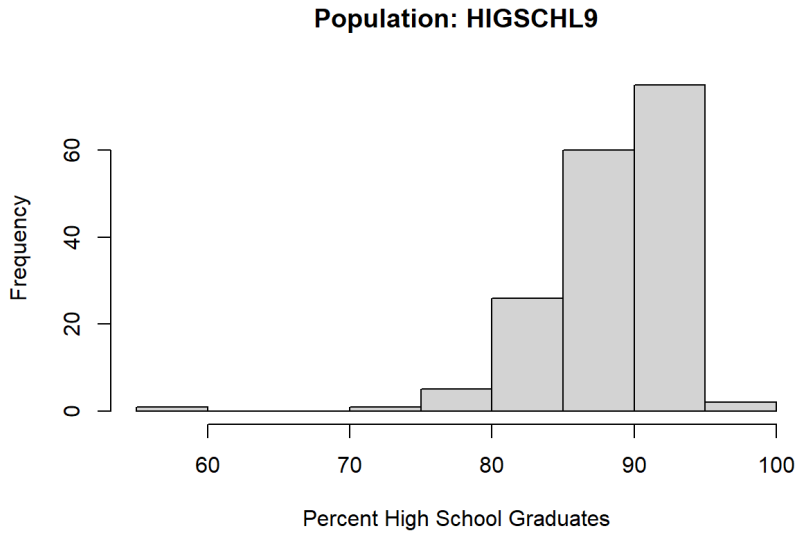
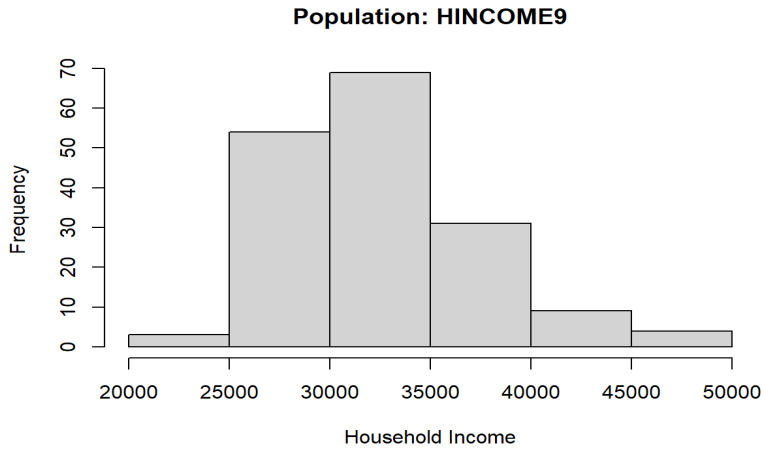
Assignment 1

1)

Variable	Mean	Median	Std Dev	Range	Skewness	Kurtosis
HIGSCHL9	88.32	89.35	4.68	36.86	-185	10.21
HINCOME9	32,418.63	31,546.45	4,864.05	26,337.04	1.11	4.67



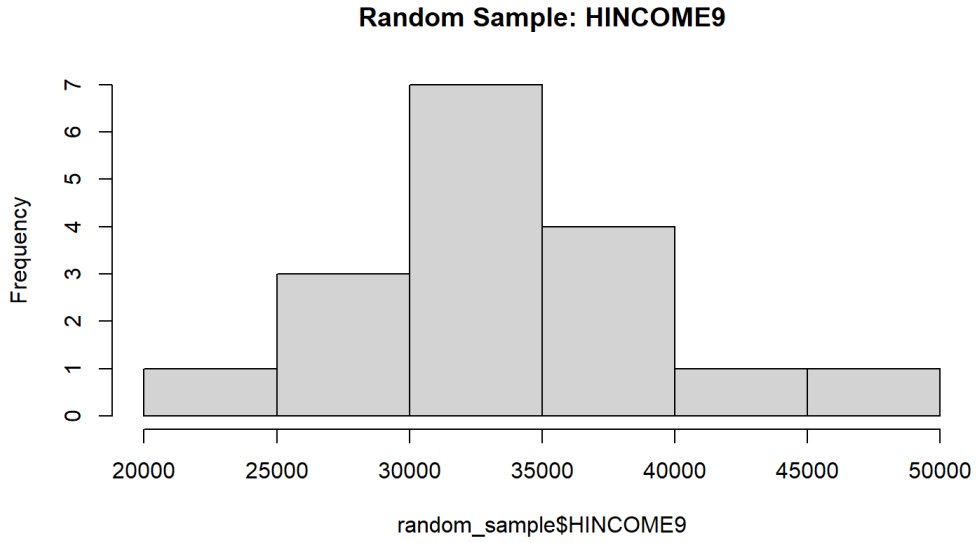


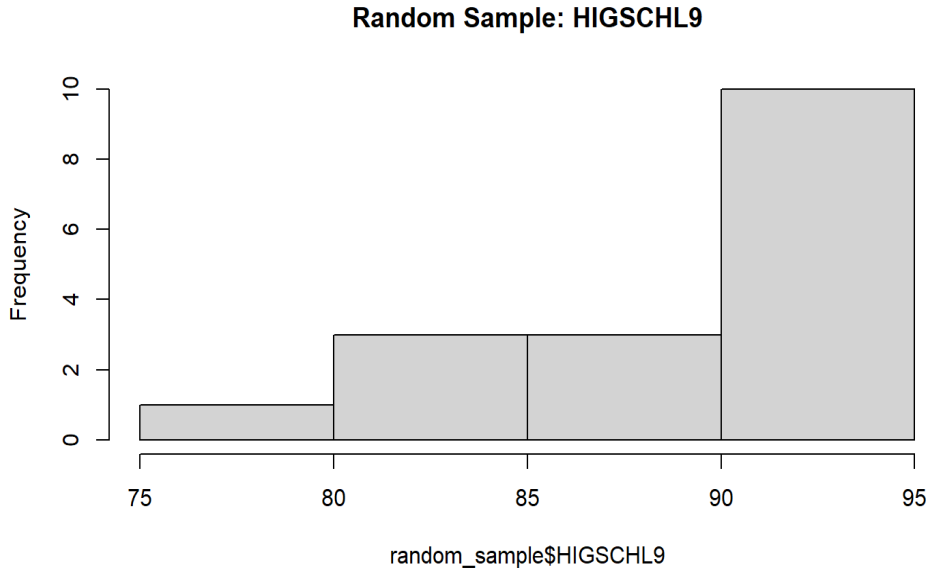
HIGSCHL9 shows high graduation rates with small variation.

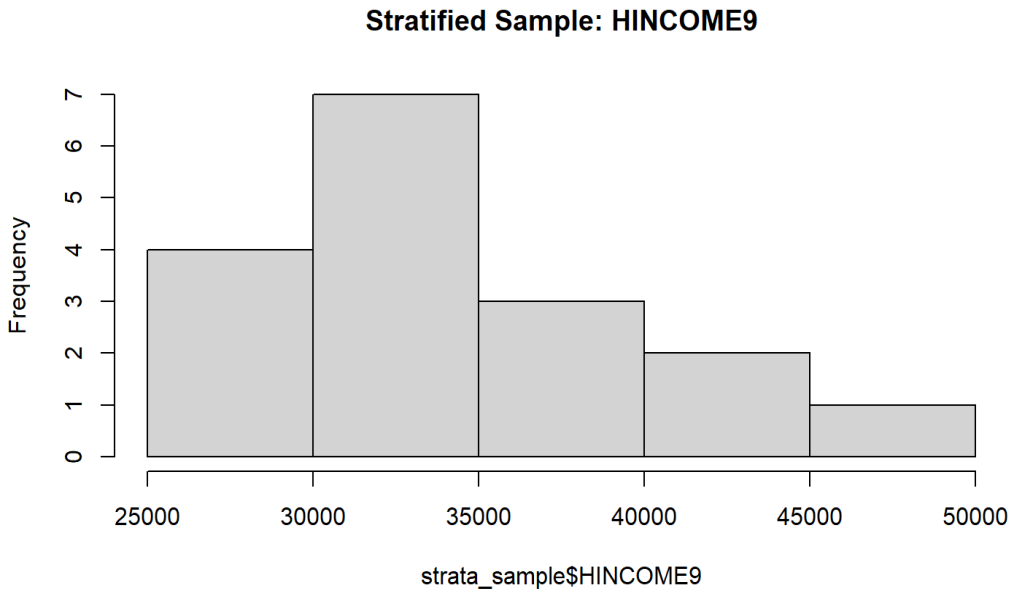
HINCOME9 shows more lower income BEAs and a few higher-income outliers.

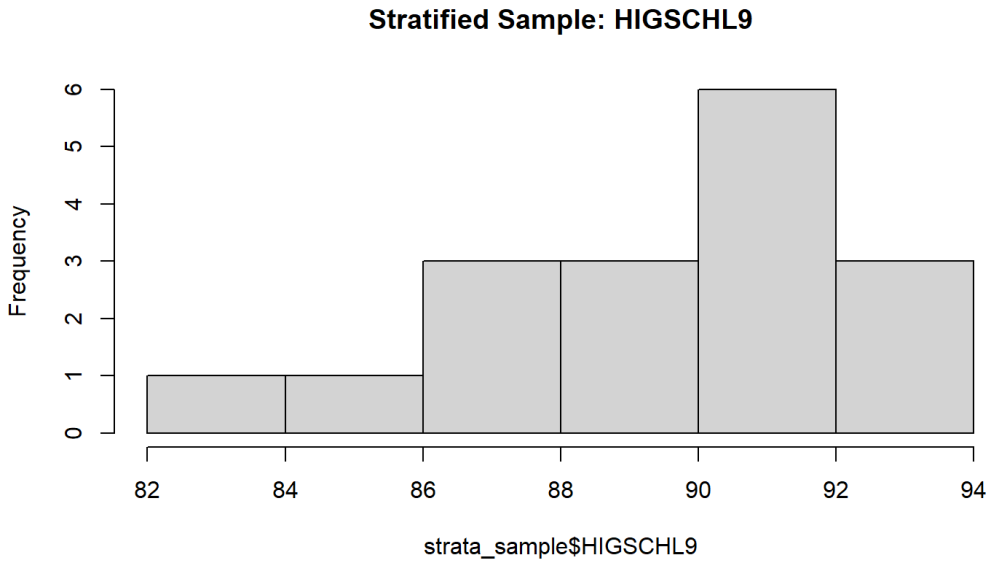
2)

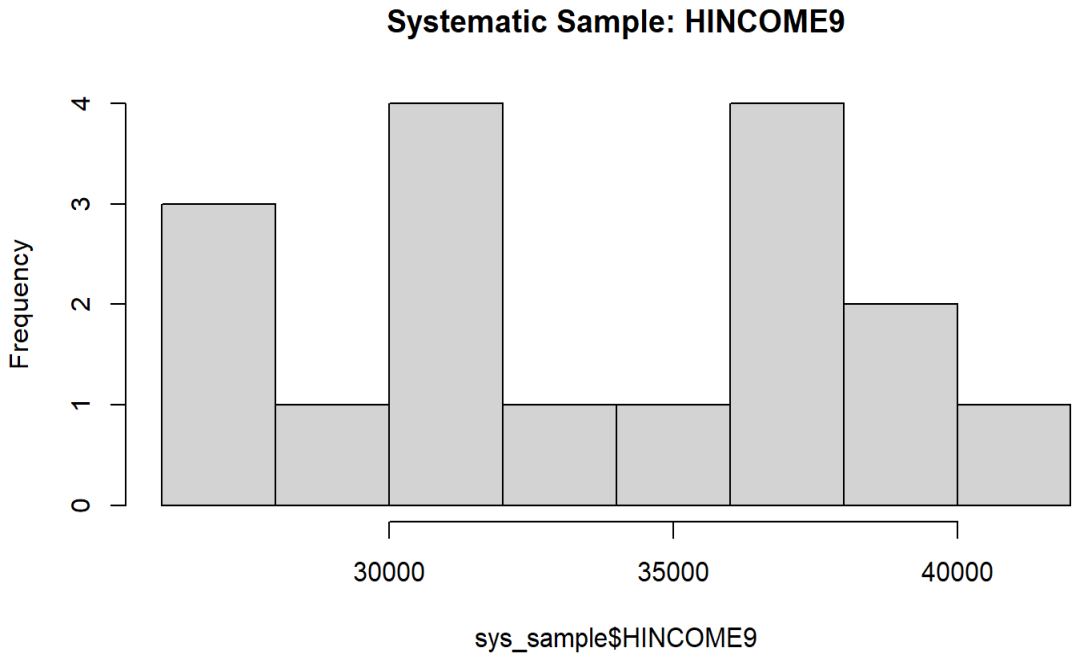
Sample	Variable	Mean	Median	Std Dev	Range	Skew	Kurt
random	HIGSCHL9	89.02	90.78	4.8	18.36	-1.31	3.91
	HINCOME9	33,602.15	31,880.01	6,406.23	26,337.04	0.91	3.75
Systematic	HIGSCHL9	90.90	92.14	3.13	10.77	-1.25	3.71
	HINCOME9	33,460.57	32,200.49	5,522.73	14,298.13	0.11	1.74
Stratified	HIGSCHL9	89.57	90.23	2.58	9.03	-0.87	2.94
	HINCOME9	34,443.01	33,139.55	5,900.33	19,334.26	1.02	2.97

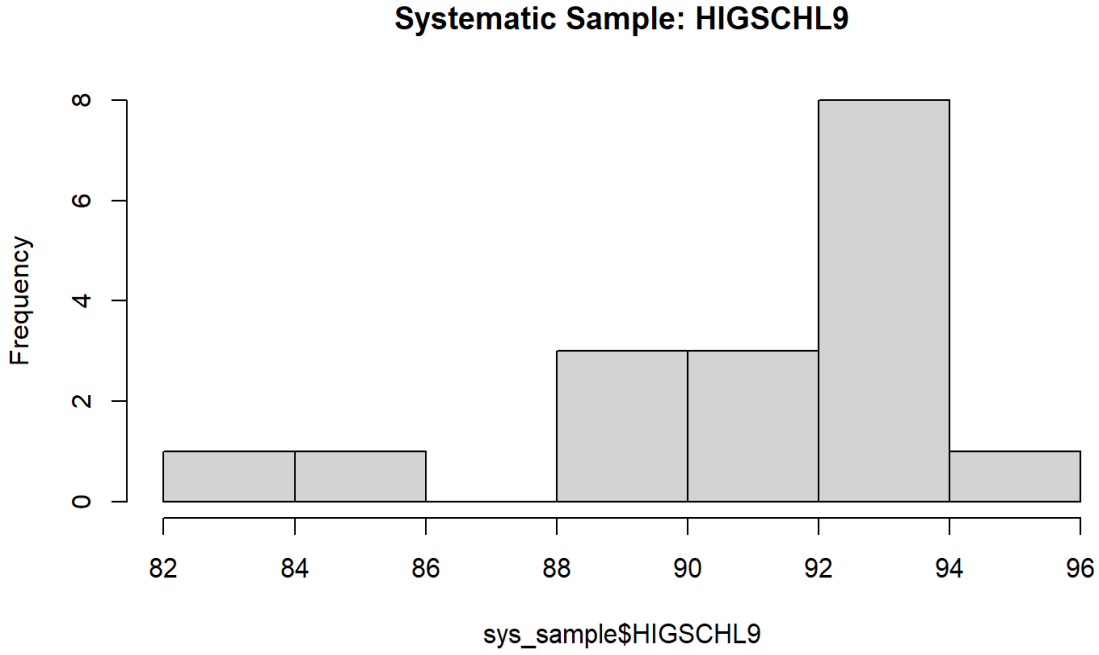





Stratified Sample: HINCOME9
Frequency
0
1
2
3
4
5
6
7
25000
30000
35000
40000
45000
50000
strata_sample$HINCOME9


Stratified Sample: HIGSCHL9
Frequency
0
1
2
3
4
5
6
82
84
86
88
90
92
94
strata_sample$HIGSCHL9


Systematic Sample: HINCOME9
Frequency
0
1
2
3
4
30000
35000
40000
sys_sample$HINCOME9


Systematic Sample: HIGSCHL9
Frequency
0
2
4
6
8
82
84
86
88
90
92
94
96
sys_sample$HIGSCHL9

3)

The stratified sample gives results most like the population. Random sampling is fine but depends on chance. Systematic can show bias if the BEA list has an internal order. Stratified reduces error by taking data from every region.

4)

There are 170 observations, and we need 17 samples, so the interval k = 10.
Picking a random start from 1 to 10 makes sure all observations have an equal chance to be chosen and keeps the selection unbiased.

5)

Proportional Stratified Sample of 30

Region	BEA Count	30 × (Count/170)	Rounded sample
Midwest	49	8.65	9
Northeast	14	2.47	3
West	31	5.47	5
South	76	13.41	13
Total	170	30	30

The BEA dataset was divided into four census regions. A proportional stratified sample of 30 was chosen using each region's share of the total population. Based on their proportions, the sample contained 9 Midwest, 2 Northeast, 5 West, and 14 South observations.

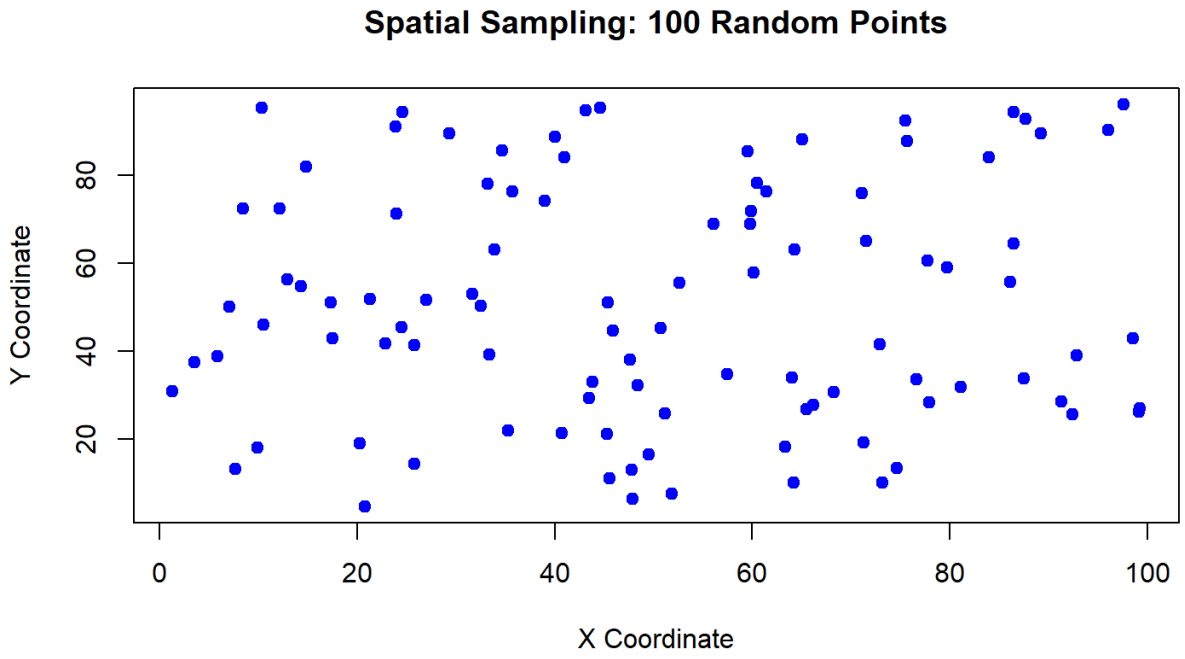
This method follows the proportional stratified sampling procedure discussed in class. Each region's sample size is calculated as:

$$n_h = n \times \frac{N_h}{N}$$

Where n is 30 the total sample, Nh is the region count, and N is 170.
Rounding was applied so that the final total equals 30. The proportional approach keeps the same regional balance as the population and helps reduce sampling error compared to taking one simple random sample.

Assignment 2)

1)



2)

Mean center: (50.29, 50.82)

The mean center is near the middle of the 0–100 range on both axes, which is what we expect for points drawn uniformly at random.

3)

Sampling approach: Simple Random Spatial Sampling.
Each point's X and Y were generated independently with a uniform distribution on 0–100, so every location in the square had the same chance of being chosen.