

# Housing\_project-DANA-4810

Khushank Sethi (100423378)

2024-10-25

```
housing <- read.csv("House_listing.csv")
attach(housing)
head(housing)
```

##	City	Price	Address	Number_Beds	Number_Baths
## 1	Trois-Rivieres	389,000	1735 rue de Bruxelles	3	2
## 2	Trois-Rivieres	865,000	1475-1477-1479, rue Houle	3	1
## 3	Trois-Rivieres	575,000	1365 rue Houle	6	2
## 4	Trois-Rivieres	447,500	1033 rue des Grives	4	1
## 5	Trois-Rivieres	925,000	4170 rue Notre-Dame Est	3	2
## 6	Trois-Rivieres	347,000	6980 rue de Honfleur	5	2
##	Province	Population	Latitude	Longitude	Median_Family_Income
## 1	Quebec	139163	46.35	-72.55	62400
## 2	Quebec	139163	46.35	-72.55	62400
## 3	Quebec	139163	46.35	-72.55	62400
## 4	Quebec	139163	46.35	-72.55	62400
## 5	Quebec	139163	46.35	-72.55	62400
## 6	Quebec	139163	46.35	-72.55	62400

```
print("missing values:")
```

```
## [1] "missing values:"
```

```
for(col in unique(names(housing))) {
  cat(col, ":", length(housing[is.na(housing[[col]])]), "\n")
}
```

```
## City : 0
## Price : 0
## Address : 0
## Number_Beds : 0
## Number_Baths : 0
## Province : 0
## Population : 0
## Latitude : 0
## Longitude : 0
## Median_Family_Income : 0
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
length(housing$Latitude[!between(Latitude, 42, 83)])
```

```
## [1] 0
```

```
length(housing$Longitude[!between(housing$Longitude, -141, -52)])
```

```
## [1] 91
```

```
housing$Longitude[housing$Longitude == 63.1005] <- -63.1005
```

```
length(housing$Longitude[!between(housing$Longitude, -141, -52)])
```

```
## [1] 0
```

```
write.csv(housing, file = "cleaned_housing.csv")
```

```
housing$Price <- as.numeric(gsub(",", "", housing$Price))
```

```
num_cols <- housing[sapply(housing, is.numeric)]
```

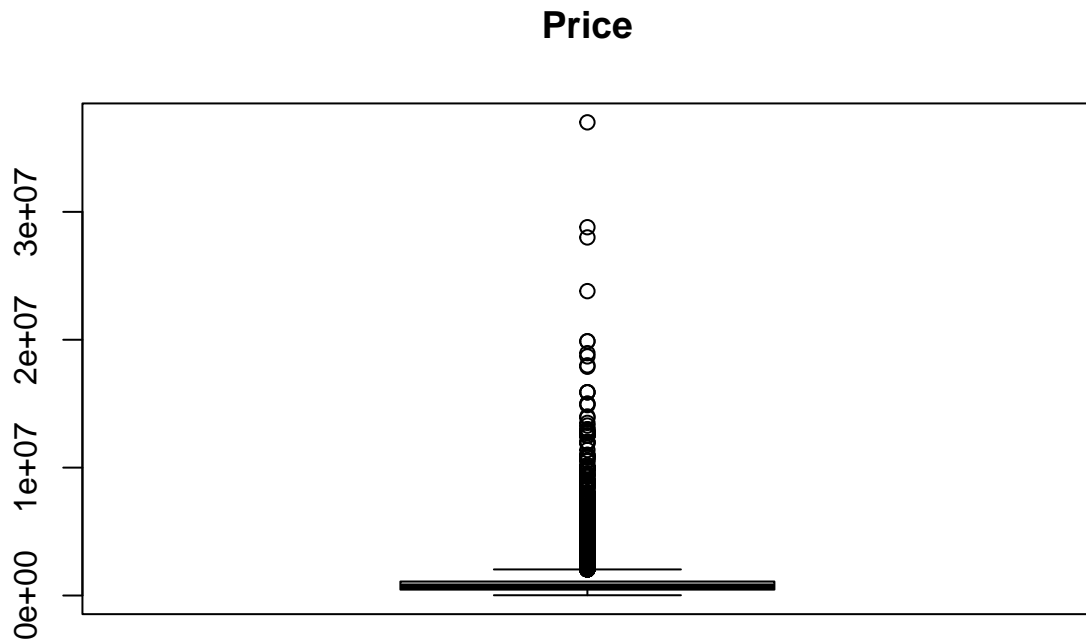
```
for(col in names(num_cols)) {
```

```
  boxplot(housing[col], main=col)
```

```
  cat(col, "outliers:")
```

```
  print(unique(boxplot.stats(housing[[col]])$out))
```

```
}
```



```
## Price outliers: [1] 2199000 2100000 2248000 2250000 2518000 2399900 10200000 2588000
## [9] 6888000 2599000 4499000 18700000 4098000 3898000 2488800 2149000
## [17] 2330000 2950000 2299800 3098000 2088000 3290000 2399000 2649999
## [25] 4198000 2699000 2898888 3199000 2695000 2960000 3200000 2499000
## [33] 7588000 2520000 2980000 3360000 2625000 2390000 2750000 2888000
## [41] 2198000 28800000 2188000 2488000 4780000 5200000 2678900 3288000
## [49] 2849000 3990000 2398900 2099000 2398000 2490000 2265000 2349000
## [57] 2899999 2988800 2588888 2599999 3199500 3850000 2558888 5999000
## [65] 4998000 2889000 5898000 3088000 2558000 2949000 3488000 2050000
## [73] 2697800 2599900 2475000 3980000 3588888 2188888 3499000 2168000
## [81] 2288000 2699999 6789000 6850000 2069000 2060000 3300000 2690000
## [89] 3700000 3680000 2380000 4299000 2799000 12500000 2150000 3875000
## [97] 2638000 2145000 5988800 2985000 3699000 3688888 2649000 2360000
## [105] 3175000 2068000 3599000 2274285 2899000 2698000 4999000 2450000
## [113] 2348000 3198000 2498000 2449000 7999999 4199000 3998000 5988000
## [121] 6990000 2849900 2999000 10800000 4188000 2098000 3780000 2499900
## [129] 2749900 2349999 2928000 4200000 4500000 17999000 3580000 4990000
## [137] 6288000 5690000 2099999 9499000 4980000 2998000 9875000 6500000
## [145] 2055000 3320000 8800000 3150000 4180000 2448000 2198888 3999999
## [153] 2170000 2879999 3880000 2688000 4668000 2990000 5188000 8990000
## [161] 2538000 8600000 2987000 5150000 2499888 2289888 2049900 2249000
## [169] 2448800 2589000 2350000 2680000 9990000 4288000 3399000 3788000
## [177] 2180000 3450000 2149999 4988000 2298000 3398888 2788000 3280000
## [185] 2995000 2898000 5498000 2880000 4380000 5998900 3888000 12488000
## [193] 2388000 6250000 2630000 8288000 6688000 3298000 3490000 4975000
## [201] 7088800 4798000 4998800 3776000 12000000 6999000 2299999 4680000
```

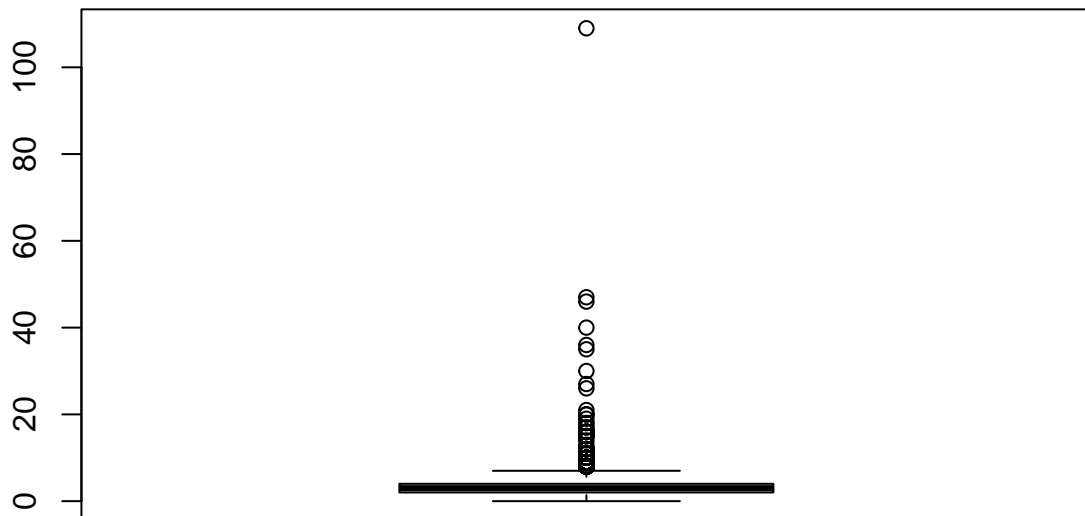
##	[209]	3799000	5980000	3195000	12799000	5288000	4300000	3799999	6800000
##	[217]	3988000	3588000	6998600	3498800	2280000	3099000	6998800	2325000
##	[225]	3080000	6899900	6380000	4100000	7999000	2299888	7500000	5590000
##	[233]	4398000	2105000	2788800	4495000	4480000	2999999	5499900	9800000
##	[241]	6650000	4088000	2500000	2299000	3795000	2650000	5998000	2895000
##	[249]	5699999	3380000	2554900	2495000	2290000	17890000	11880000	2780000
##	[257]	2275000	5500000	3500000	7498000	4088888	3388000	27999999	2698800
##	[265]	8680000	8550000	4888800	2478888	12880000	5800000	14888000	3800000
##	[273]	2600000	7800000	2300000	3180000	2700000	3250000	8999999	10999999
##	[281]	14000000	7900000	3999900	4600000	2849999	3900000	2200000	2347000
##	[289]	4400000	2198800	8999900	4199900	2190000	2550000	2595000	2499999
##	[297]	3695000	2199900	3299900	2188800	2175000	2149900	2545000	2215990
##	[305]	3299000	4750000	2699900	2225000	2432000	4900000	2249500	2299900
##	[313]	2195000	7360000	2274999	3658000	3750000	3000000	5499000	3295000
##	[321]	8500000	2420000	3988888	2470900	2295000	5599000	2058000	2399999
##	[329]	2988000	3100000	2138000	5198000	3348888	2298800	4099999	2400000
##	[337]	3188000	2628000	2850000	2598800	2480000	9499900	3568000	2496000
##	[345]	2678800	2438000	15000000	3688000	2999888	3288800	3790000	2169888
##	[353]	2070000	2099800	3650000	3090000	2090000	2689000	2238000	2879000
##	[361]	2078000	2080000	2898900	3168000	2938000	2499800	2089000	4390800
##	[369]	3590000	6698800	3498000	2288800	2999800	2229879	2365000	3449800
##	[377]	2568000	2072400	2099900	2068888	2349900	8900000	3499900	2549000
##	[385]	3995000	3050000	6295000	2598000	7880000	6370000	4488000	4280000
##	[393]	12698000	10000000	9980000	3598000	2590000	2148000	5278000	4580000
##	[401]	19880000	2890000	2489000	2159900	4250000	2570000	9490000	2724900
##	[409]	2339900	5880000	2789000	6580000	3438000	2580000	2229000	4498000
##	[417]	2649900	2665900	2888888	2439900	4988800	5950000	7699900	3190000
##	[425]	2199888	4375000	2745000	4199500	3425000	3350000	5000000	2875000
##	[433]	2099888	3690000	3999987	5559000	2636000	2800000	2375000	7299980
##	[441]	4150000	2296000	2799900	3240000	7688000	9380000	6499000	2160000
##	[449]	5380000	3482000	3950000	2249900	4249900	2999995	7750000	2790000
##	[457]	8299000	3549999	5495000	5388000	4799000	4890000	2125000	2679000
##	[465]	2245000	3895000	4995000	2425000	2395000	2174900	2458000	2185000
##	[473]	4999900	3258000	2795000	3575000	4650000	2189000	6700000	2999900
##	[481]	2675000	2900000	7199000	2368000	3999000	4800000	6495000	3400000
##	[489]	3399900	5700000	2260000	5790000	6995000	3699900	3749000	3480000
##	[497]	4888000	3279900	4298000	2797000	5780000	3659000	2049000	3949000
##	[505]	7680000	2178000	4950000	7995000	4089000	5375000	6788000	4899000
##	[513]	6697000	4700000	9998000	18950000	13288000	5899000	10880000	23800000
##	[521]	3238000	8998000	10780000	7380000	3188800	4390000	6980000	4688000
##	[529]	4175000	15888000	7998000	4498888	2798000	4295000	6698000	3498888
##	[537]	3475000	3798000	4788000	10080000	11380000	4388000	5598000	6388000
##	[545]	6680000	3785000	6080000	4399000	5680000	2749000	2998800	8888000
##	[553]	9280000	8980000	2575000	2088800	2338000	4049800	4880000	3550000
##	[561]	10995000	2249999	7299000	4499900	4350000	6695000	3195777	4450000
##	[569]	2288888	2379900	2870000	2199999	3372000	5600000	3395000	5999999
##	[577]	5400000	10600000	2359900	2777700	4325000	2349888	8888888	2479999
##	[585]	2295900	4690000	4299750	7600000	3160000	2140000	3448800	2798888
##	[593]	6600000	4249000	4788888	9950000	6900000	3749900	2075000	4745000
##	[601]	3488888	2924900	2199990	2559000	2056000	2899900	5499999	13499999
##	[609]	3077777	2049999	2192000	2335000	3275000	4463445	2315000	2388800
##	[617]	3679000	5678000	8478000	2095000	2488888	3495000	2588168	2665000
##	[625]	2379000	12750000	3349000	8688000	3849000	4336900	3249000	2285000
##	[633]	2159000	2849888	3339000	3599900	4739000	3199999	2499998	6095000

```

## [641] 3379000 3149900 2588800 7595000 3999990 3369000 3595000 5825000
## [649] 2688888 2498888 5788000 4289000 5588000 3726000 2298999 7888000
## [657] 5750000 3598888 10950000 2189999 3249900 2460000 2139800 7495000
## [665] 2099600 2127000 3600000 3200024 6290000 13900000 11990000 37000000
## [673] 2749999 3149000 2345000 2739000 6395000 3399999 2789900 2805000
## [681] 2479000 2345900 5300000 2499990 2199800 2147382 2998900 2149500
## [689] 10500000 4190000 5915100 2469000 6300000 2215000 2289000 14998600
## [697] 2385000 3375000 3499800 2579000 4488800 3148888 2394000 2431408
## [705] 12999888 2388888 4850000 6000000 3889000 4999990 4699000 3698800
## [713] 2098600 2279990 2998600 2448888 15900000 5888888 8488000 5280000
## [721] 2379999 5900000 3449000 4888777 9000000 2279000 3698000 13000000
## [729] 2248888 5250000 2249888 7980000 8000000 3298800 2298888 3177700
## [737] 5199000 3499999 2899888 7950000 8999000 2888800 3587000 5549000
## [745] 2728900 2948000 3498900 4599990 4999880 6950000 2197000 9995000
## [753] 3398000 5399999 2799999 3777000 5595000 2397000 3199900 2268800
## [761] 2179000 2098800 5849000 9888888 5990000 2800800 3350333 7899900
## [769] 2399888 2278000 6745000 3549900 3749800 3999711 5749000 5995000
## [777] 2825000 2299990

```

**Number\_Beds**

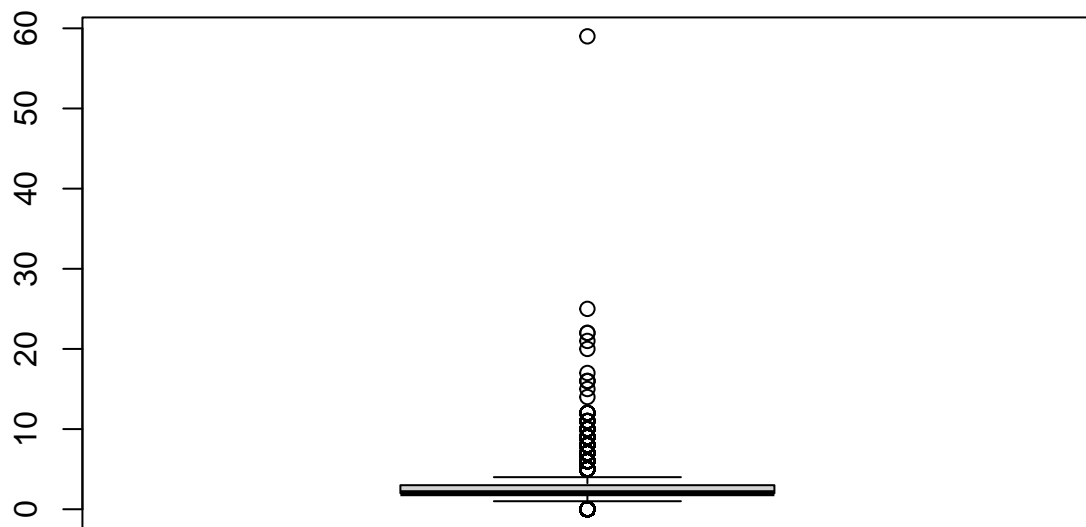


```

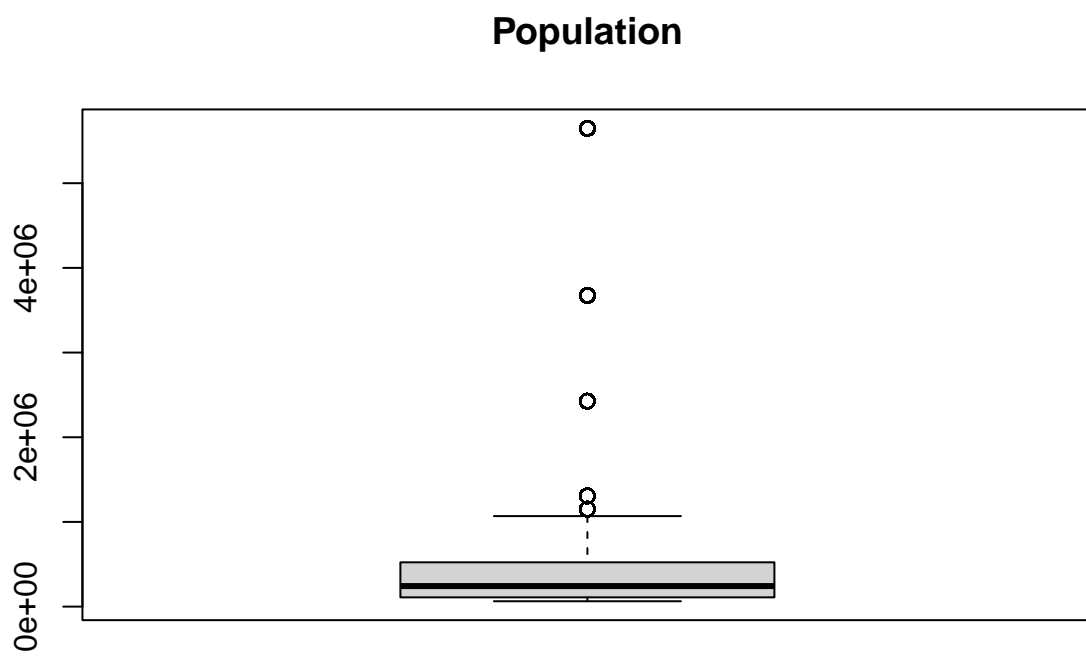
## Number_Beds outliers: [1] 12 9 8 11 18 10 13 19 21 15 17 35 16 20 27 26 40 30 3
## [20] 14 46 109 47

```

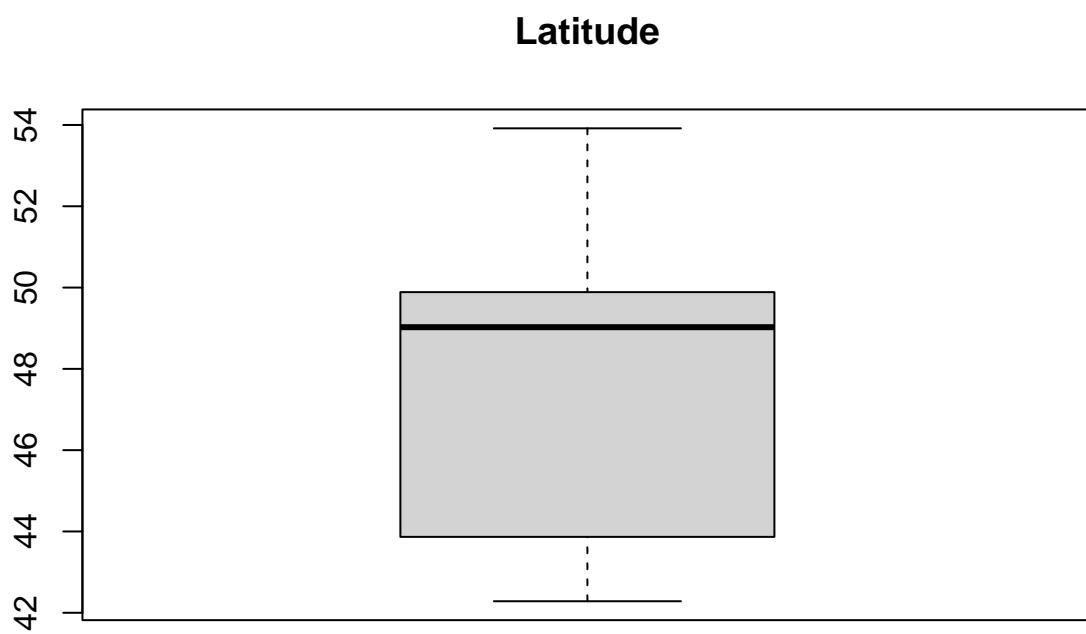
## Number\_Baths



```
## Number_Baths outliers: [1] 12 7 6 5 0 8 10 9 11 21 15 20 14 22 16 25 17 59
```

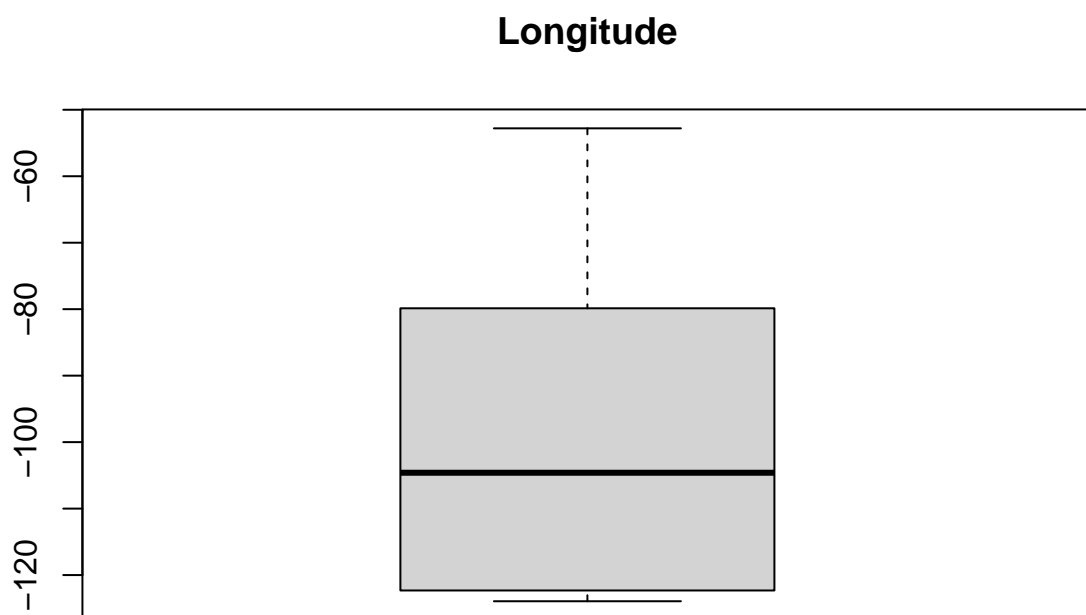


```
## Population outliers:[1] 3675219 2426160 1151635 5647656 1306784
```



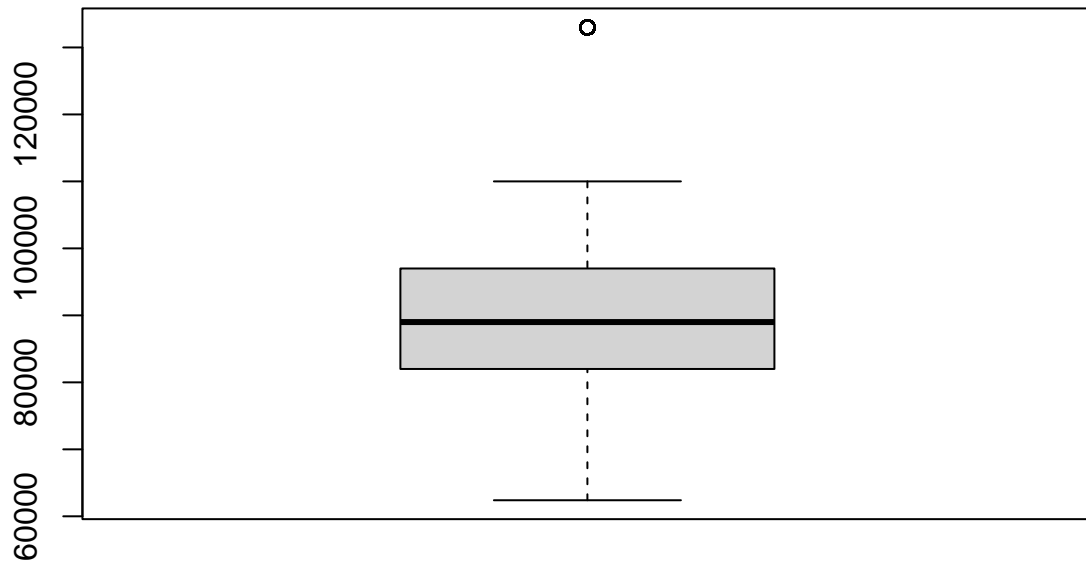
```
## Latitude outliers:numeric(0)
```





```
## Longitude outliers:numeric(0)
```

## Median\_Family\_Income



```
## Median_Family_Income outliers:[1] 133000
```