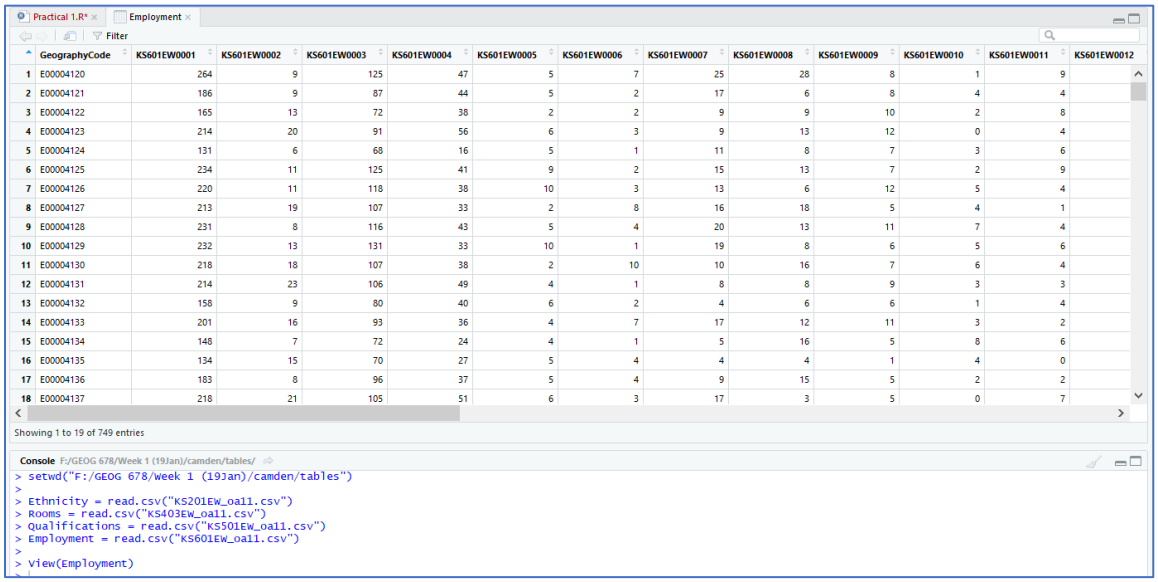
Assignment 1: Spatial Analysis and Viz with R (I)

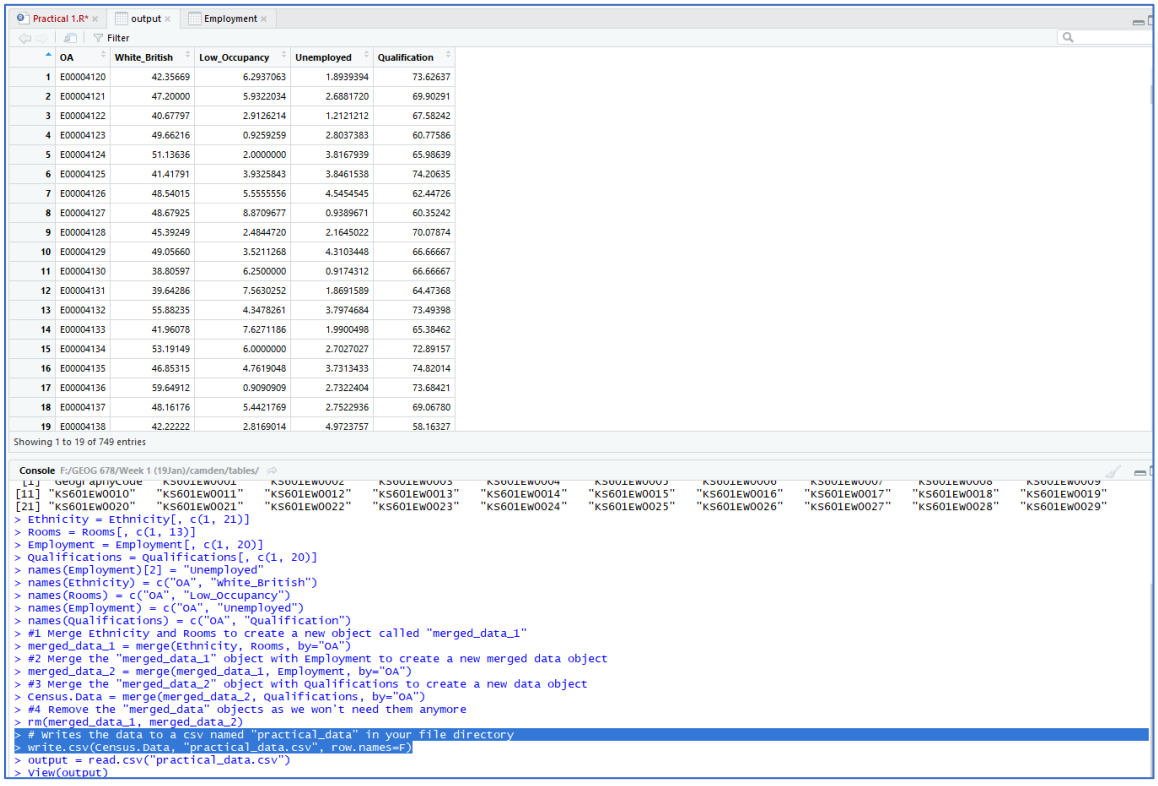
1: Screenshots of at least 2 critical results from each practical exercise. Please label each one properly.

Practical 1:



	GeographyCode	KS601EW0001	KS601EW0002	KS601EW0003	KS601EW0004	KS601EW0005	KS601EW0006	KS601EW0007	KS601EW0008	KS601EW0009	KS601EW0010	KS601EW0011	KS601EW0012
1	E00004120	264	9	125	47	5	7	25	28	8	1	9	
2	E00004121	186	9	87	44	5	2	17	6	8	4	4	
3	E00004122	165	13	72	38	2	2	9	9	10	2	8	
4	E00004123	214	20	91	56	6	3	9	13	12	0	4	
5	E00004124	131	6	68	16	5	1	11	8	7	3	6	
6	E00004125	234	11	125	41	9	2	15	13	7	2	9	
7	E00004126	220	11	118	38	10	3	13	6	12	5	4	
8	E00004127	213	19	107	33	2	8	16	18	5	4	1	
9	E00004128	231	8	116	43	5	4	20	13	11	7	4	
10	E00004129	232	13	131	33	10	1	19	8	6	5	6	
11	E00004130	218	18	107	38	2	10	10	16	7	6	4	
12	E00004131	214	23	106	49	4	1	8	8	9	3	3	
13	E00004132	158	9	80	40	6	2	4	6	6	1	4	
14	E00004133	201	16	93	36	4	7	17	12	11	3	2	
15	E00004134	148	7	72	24	4	1	5	16	5	8	6	
16	E00004135	134	15	70	27	5	4	4	4	1	4	0	
17	E00004136	183	8	96	37	5	4	9	15	5	2	2	
18	E00004137	218	21	105	51	6	3	17	3	5	0	7	

Figure 1. Data in csv file viewed using View() function.



	OA	White_British	Low_Occupancy	Unemployed	Qualification
1	E00004120	42.35669	6.2937063	1.8939394	73.62637
2	E00004121	47.20000	5.9322034	2.6881720	69.90291
3	E00004122	40.67797	2.9126214	1.2121212	67.58242
4	E00004123	49.66216	0.9259259	2.8037383	60.77586
5	E00004124	51.13636	2.0000000	3.8167939	65.98639
6	E00004125	41.41791	3.9325843	3.8461538	74.20635
7	E00004126	48.54015	5.5555556	4.5454545	62.44726
8	E00004127	48.67925	8.8709677	0.9389671	60.35242
9	E00004128	45.39249	2.4844720	2.1645022	70.07874
10	E00004129	49.05660	3.5211268	4.3103448	66.66667
11	E00004130	38.80597	6.2500000	0.9174312	66.66667
12	E00004131	39.64286	7.5630252	1.8691589	64.47368
13	E00004132	55.88235	4.3478261	3.7974684	73.49398
14	E00004133	41.96078	7.6271186	1.9900498	65.38462
15	E00004134	53.19149	6.0000000	2.7027027	72.89157
16	E00004135	46.85315	4.7619048	3.7313433	74.82014
17	E00004136	59.64912	0.9090909	2.7322404	73.68421
18	E00004137	48.16176	5.4421769	2.7522936	69.06780
19	E00004138	42.22222	2.8169014	4.9723757	58.16327

Figure 2. Data from Census.Data exported to practical_data csv file using write.csv

Practical 2:

```
> mean(Census.Data$Unemployed)
[1] 4.510309
> median(Census.Data$Unemployed)
[1] 4.186047
> range(Census.Data$Unemployed)
[1]  0.00000 18.62348
> summary(Census.Data)
      OA              White_British    Low_Occupancy      Unemployed     Qualification
 Length:749         Min.   : 7.882   Min.   : 0.000   Min.   : 0.000   Min.   :11.64
 Class :character   1st Qu.:35.915   1st Qu.: 6.015   1st Qu.: 2.500   1st Qu.:36.32
 Mode  :character   Median :44.541   Median :10.000   Median : 4.186   Median :55.10
                    Mean   :44.832   Mean   :11.597   Mean   : 4.510   Mean   :51.43
                    3rd Qu.:54.472   3rd Qu.:16.107   3rd Qu.: 6.158   3rd Qu.:66.23
                    Max.   :78.035   Max.   :64.286   Max.   :18.623   Max.   :88.07
```

Figure 3. Descriptive statistics using functions such as mean(), median(), range() and summary().

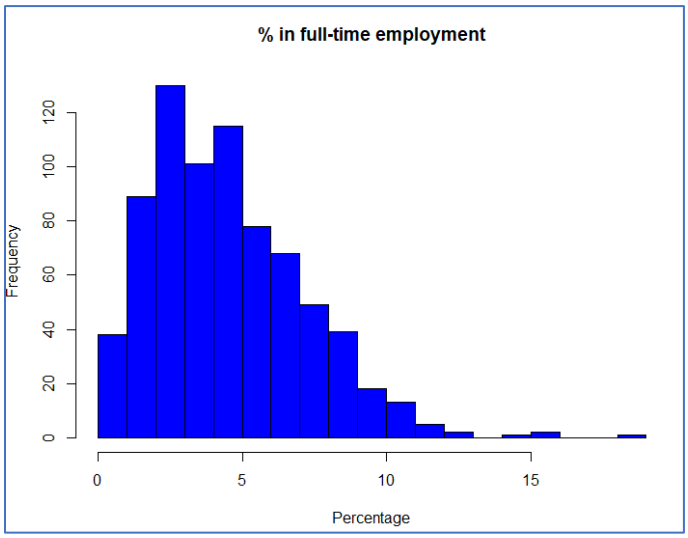


Figure 4. Histogram of percentage of full-time employment.

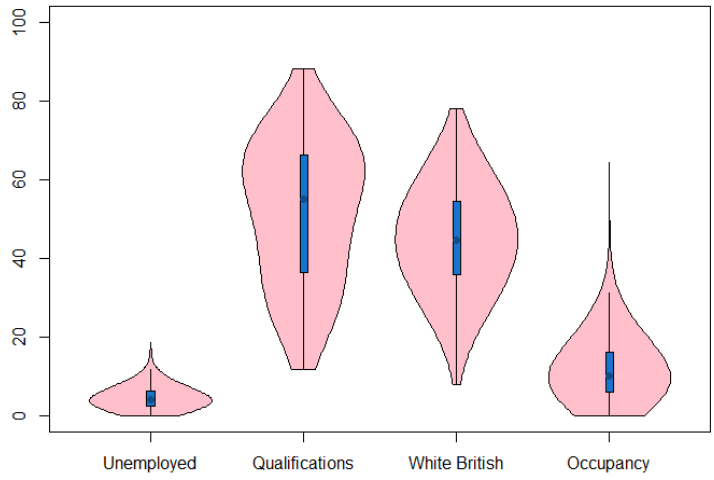


Figure 5. Violin plot for 4 variables – Unemployment, Qualifications, White British and Occupancy.

Assignment 1: Spatial Analysis and Viz with R (I)

Practical 3:

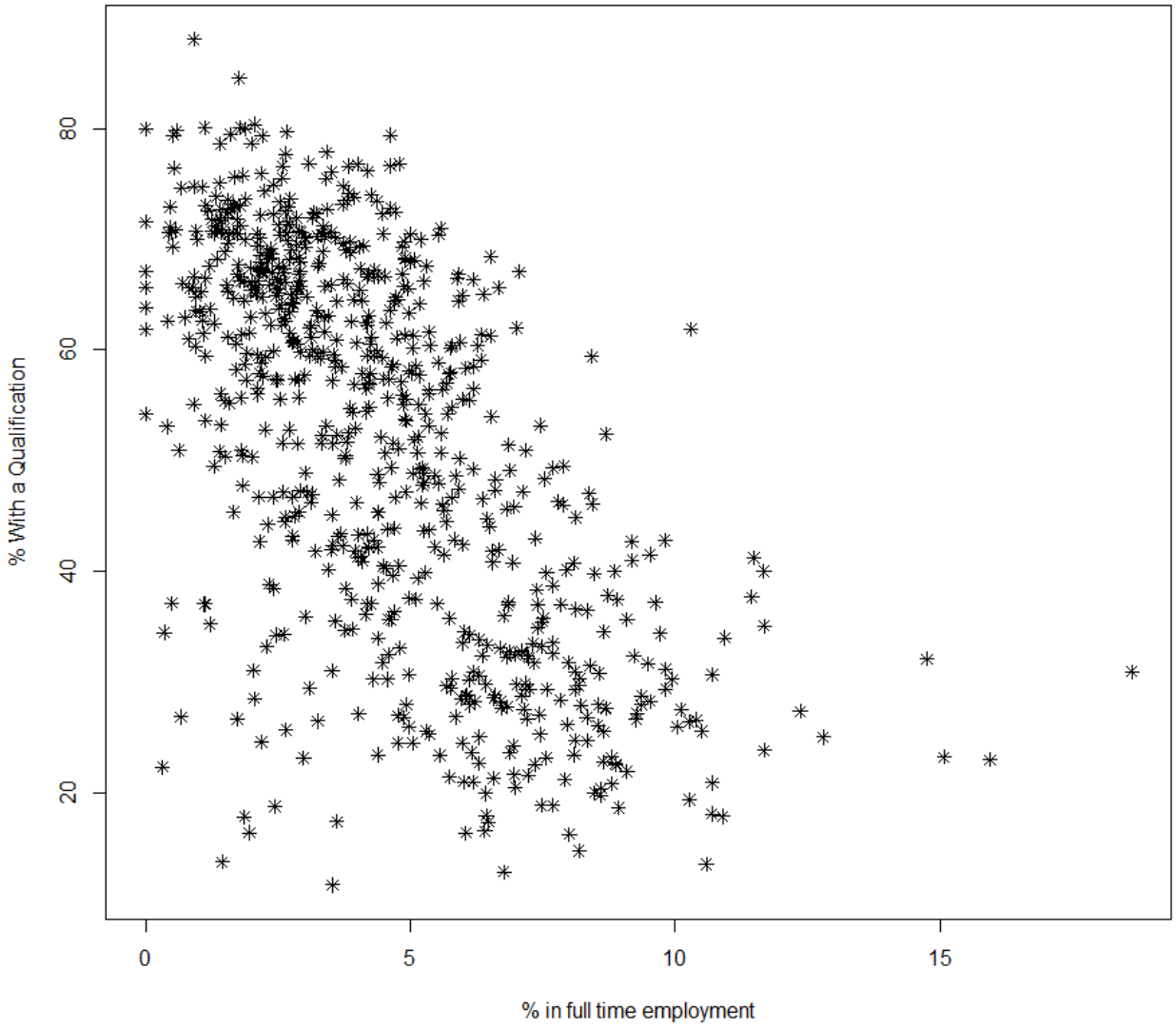


Figure 6. Simple scatter plot between percentages of full-time employment and qualification, using plot() function.

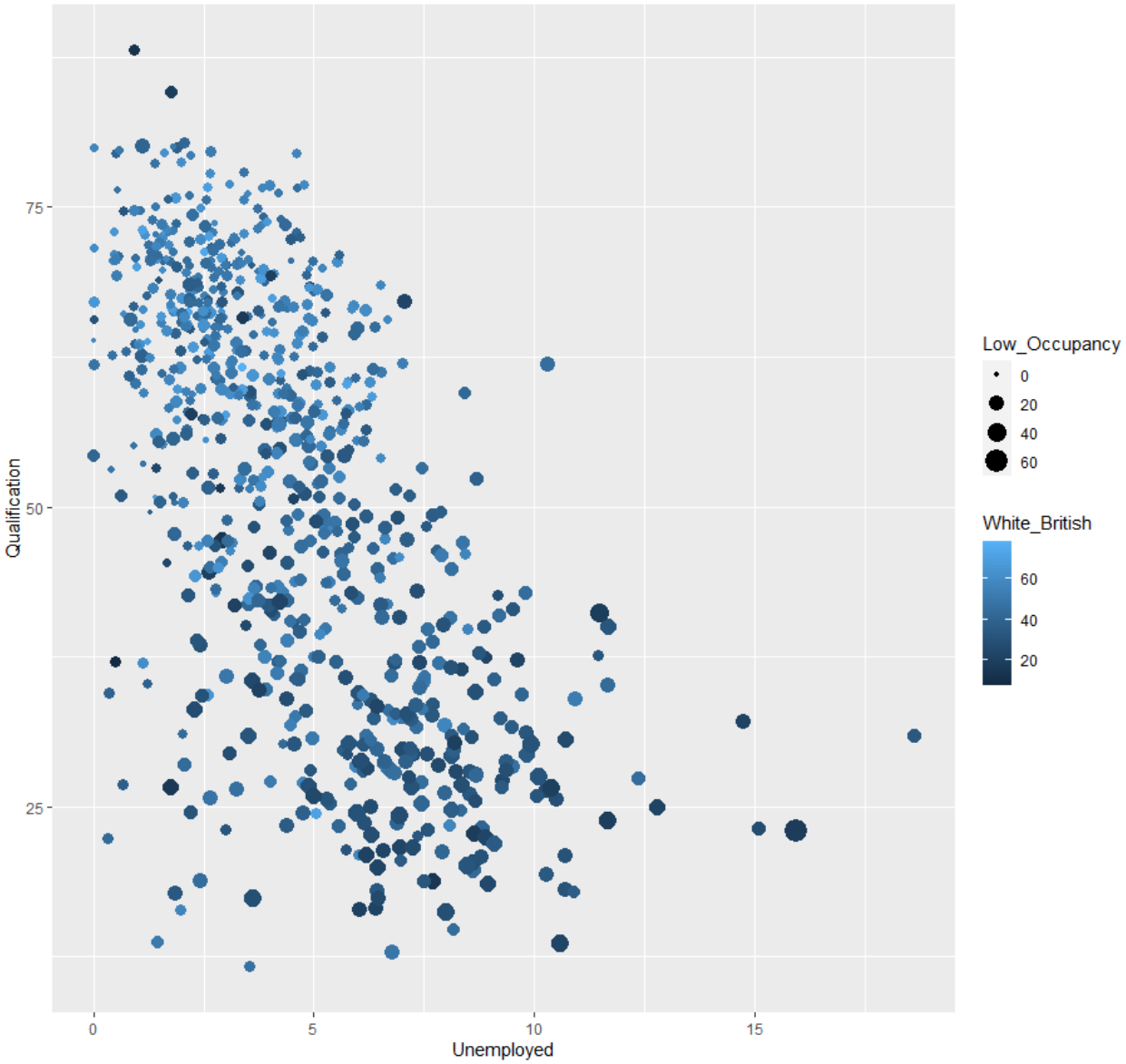


Figure 7. Data visualization using ggplot() function. Scatter plot between variables unemployed and Qualification.

Practical 4:

```
> cor.test(Census.Data$Unemployed, Census.Data$Qualification)

        Pearson's product-moment correlation

data:  Census.Data$Unemployed and Census.Data$Qualification
t = -21.85, df = 747, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.6662641 -0.5786800
sample estimates:
      cor
-0.624431

> cor.test(Census.Data$Unemployed, Census.Data$Qualification, method="spearman")

        Spearman's rank correlation rho

data:  Census.Data$Unemployed and Census.Data$Qualification
S = 113733998, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.6240406
```

Figure 8. Relationship between two variables, unemployed and qualification, using Pearson and Spearman correlation coefficients.

```
> model_1 <- lm(Census.Data$Qualification~ Census.Data$Unemployed)
> plot(Census.Data$Unemployed, Census.Data$Qualification, xlab="% Unemployed",
+      ylab="% With a Qualification") + abline (model_1)
integer(0)
> summary(model_1)

Call:
lm(formula = Census.Data$Qualification ~ Census.Data$Unemployed)

Residuals:
    Min      1Q  Median      3Q     Max
-50.172  -9.635   2.339   9.512  36.887

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)             69.7740     0.9743   71.61   <2e-16 ***
Census.Data$Unemployed  -4.0672     0.1861  -21.85   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.53 on 747 degrees of freedom
Multiple R-squared:  0.3899,    Adjusted R-squared:  0.3891
F-statistic: 477.4 on 1 and 747 DF,  p-value: < 2.2e-16
```

Figure 9. Linear regression model results obtained using lm() function.

Assignment 1: Spatial Analysis and Viz with R (I)

Practical 5:

```
> OA.Census <- merge(Output.Areas, Census.Data, by.x="OA11CD", by.y="OA")
> OA.Census
class       : SpatialPolygonsDataFrame
features    : 749
extent      : 523954.5, 531554.9, 180959.8, 187603.6  (xmin, xmax, ymin, ymax)
crs         : +proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717 +x_0=400000 +y_0=-100000 +ellps=airy +units=m +no_defs
variables   : 5
names       :    OA11CD,    White_British,    Low_Occupancy,       Unemployed,    Qualification
min values  : E00004120, 7.88177339901478,                0,                0, 11.6438356164384
max values  : E00174680, 78.0346820809249, 64.2857142857143, 18.6234817813765, 88.0733944954129
```

Figure 10. Census.Data joined to shapefile Camden_oa11 using merge() function.

```
> writeOGR(OA.Census, dsn = getwd(),
+          layer = "Census_OA_Shapefile", driver="ESRI Shapefile")
Warning message:
In writeOGR(OA.Census, dsn = getwd(), layer = "Census_OA_Shapefile",  :
  Field names abbreviated for ESRI Shapefile driver
> output = readOGR(".", "Census_OA_Shapefile")
OGR data source with driver: ESRI Shapefile
Source: "F:\GEOG 678\Week 1 (19Jan)\camden\shapefiles", layer: "Census_OA_Shapefile"
with 749 features
It has 5 fields
```

Figure 11. The Census data (OA.Census) saved as Cenus_OA_Shapefile using writeOCG() function.

Practical 6:

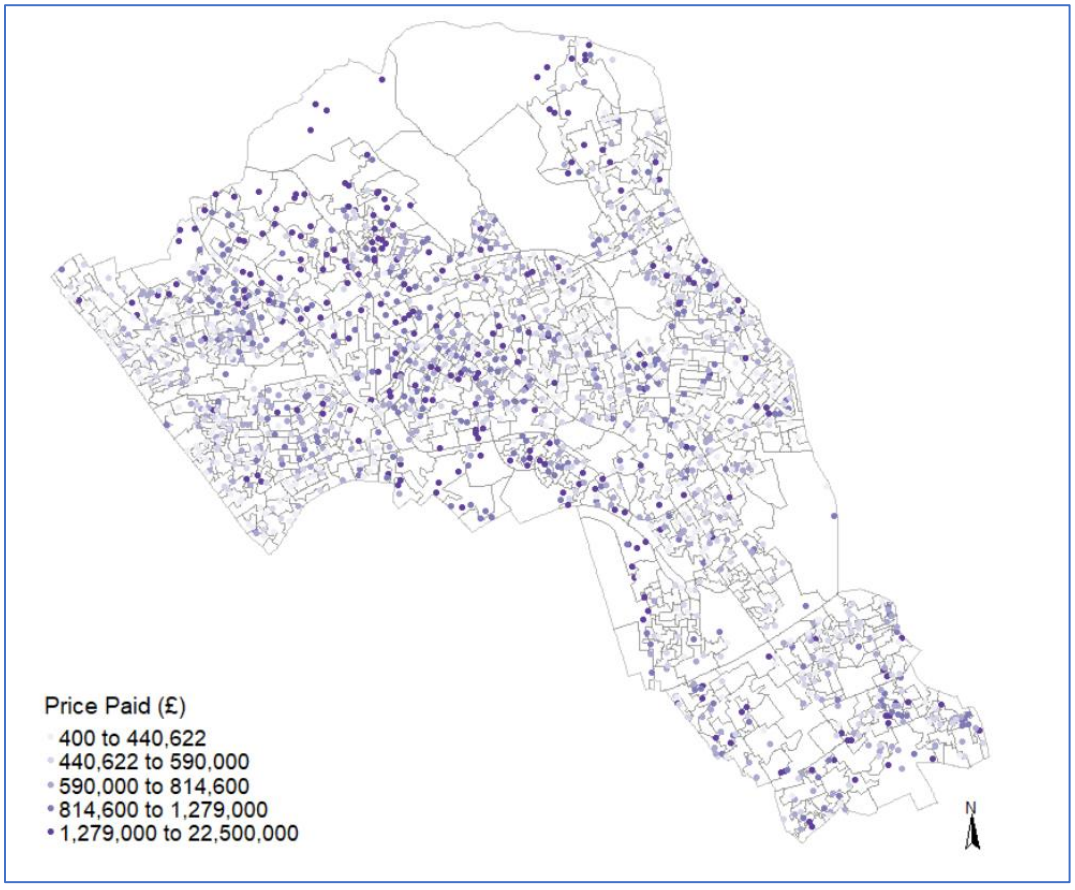


Figure 12. Colored dotted map with legend and north arrow showing house price paid in Camden, London.

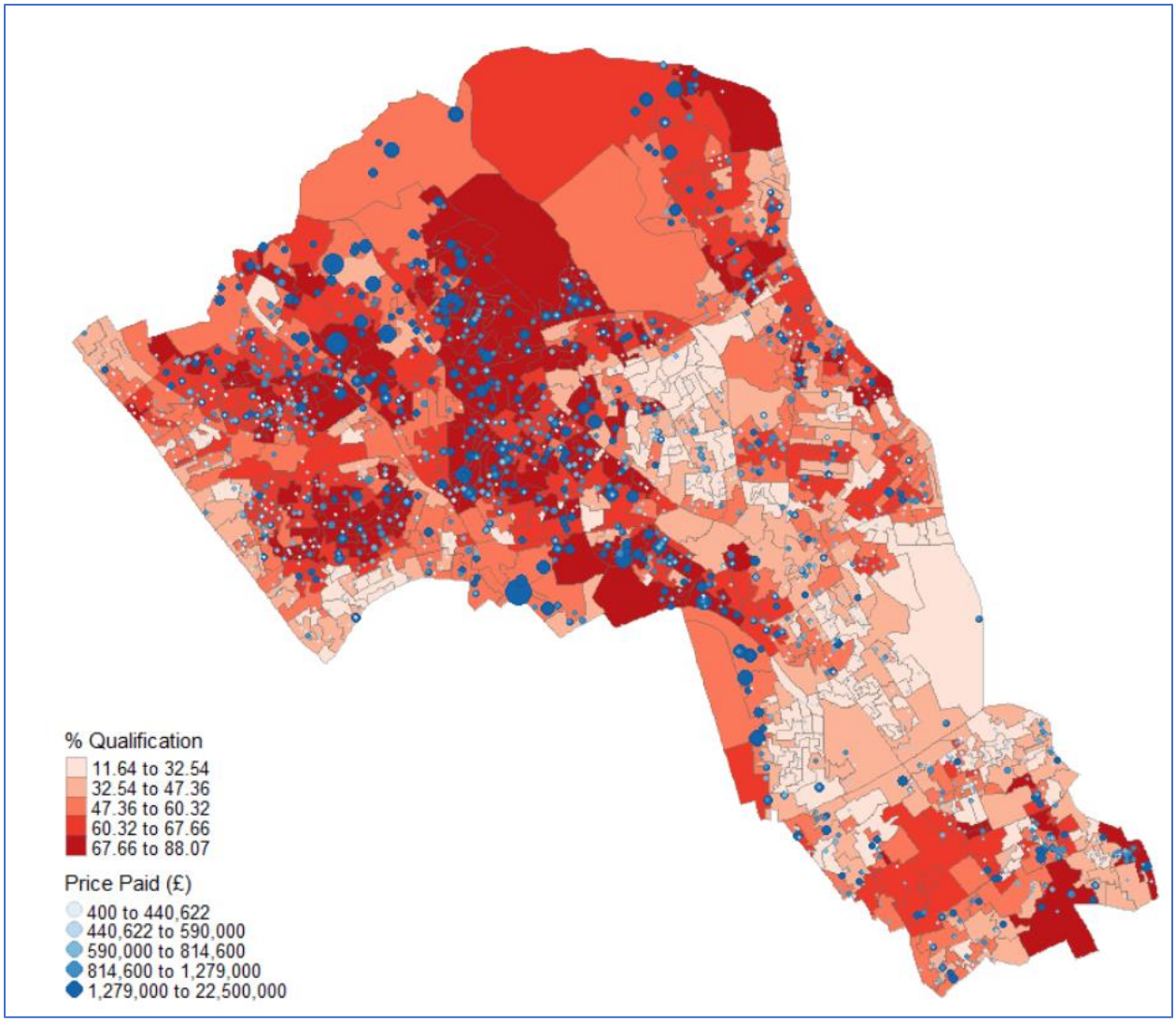


Figure 13. House price paid shown with proportionate symbols on a choropleth map of % Qualification.

2: Screenshots of all plots, graphics, and maps. Please label each one properly.

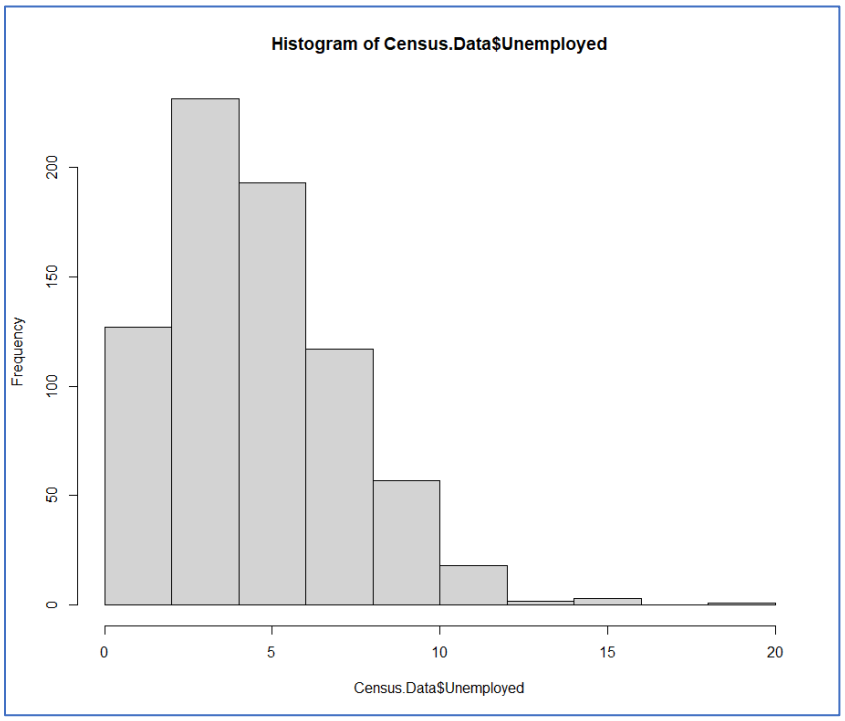


Figure 14. Histogram of the unemployed variable created using hist() function.

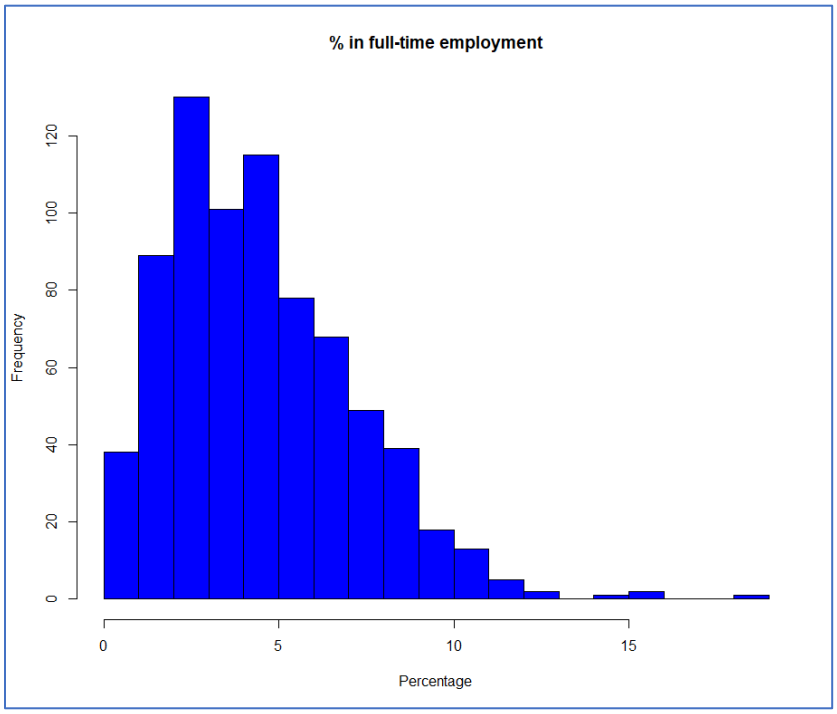


Figure 15. Enhanced histogram of the percentage in full-time employment created by supplying additional parameters to hist() function.

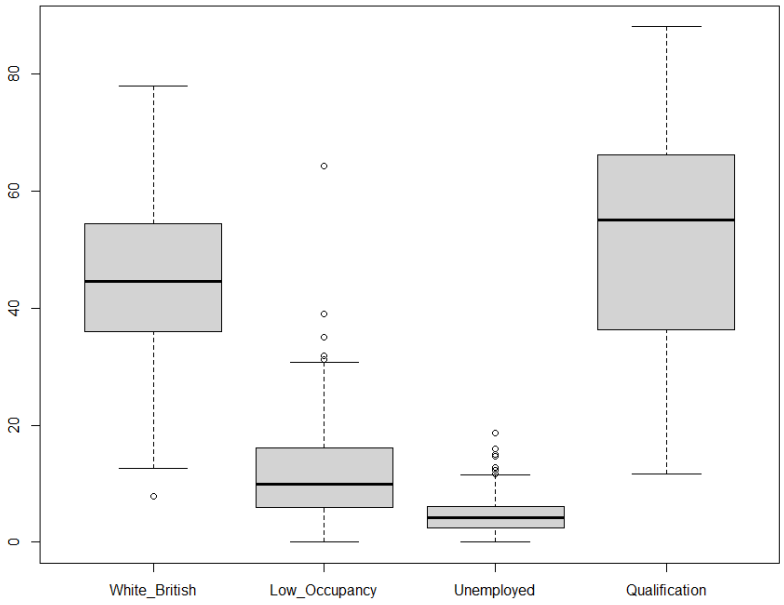


Figure 16. Box and whisker plots for various variables in Census.Data created using boxplot() function.

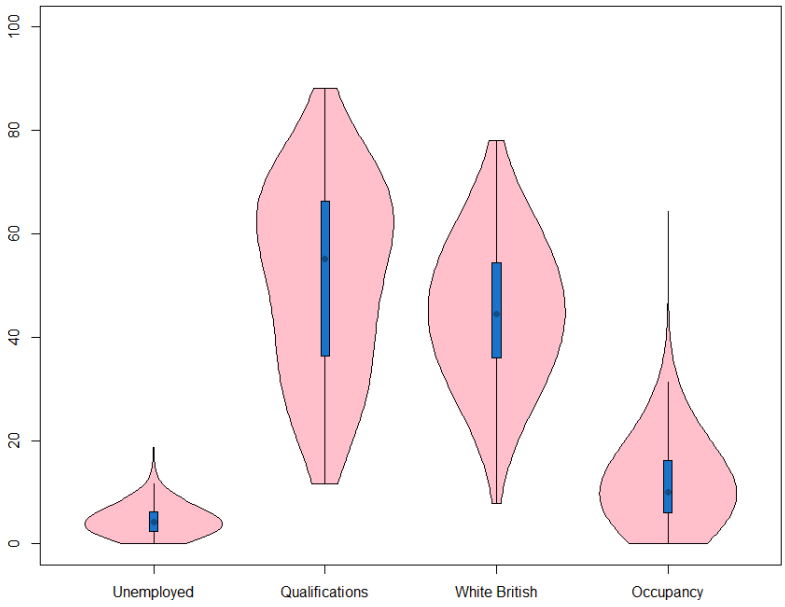


Figure 17. Violin plot for various variable in Census.Data created using vioplot() function.

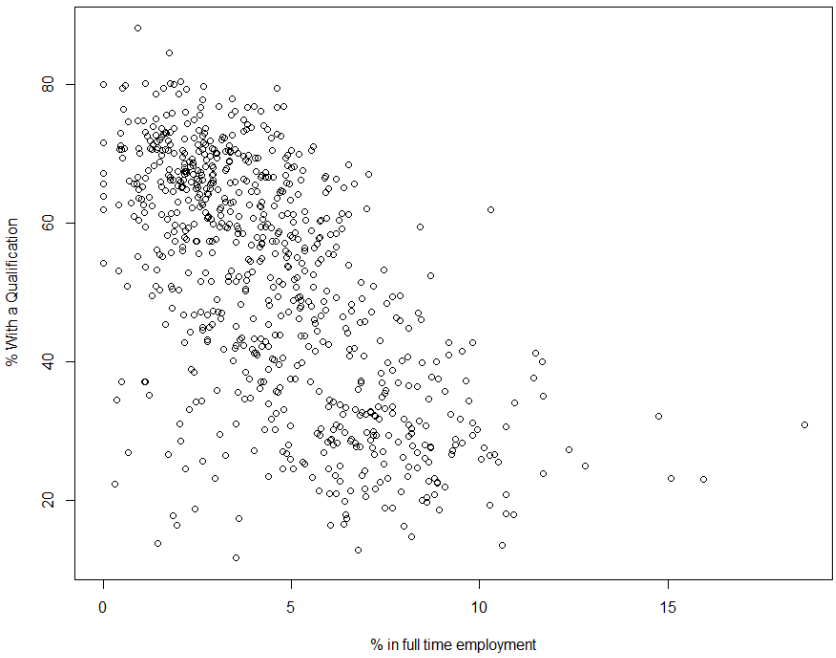


Figure 18. Simple scatter plot between two variables – employment and qualification, using plot() function.

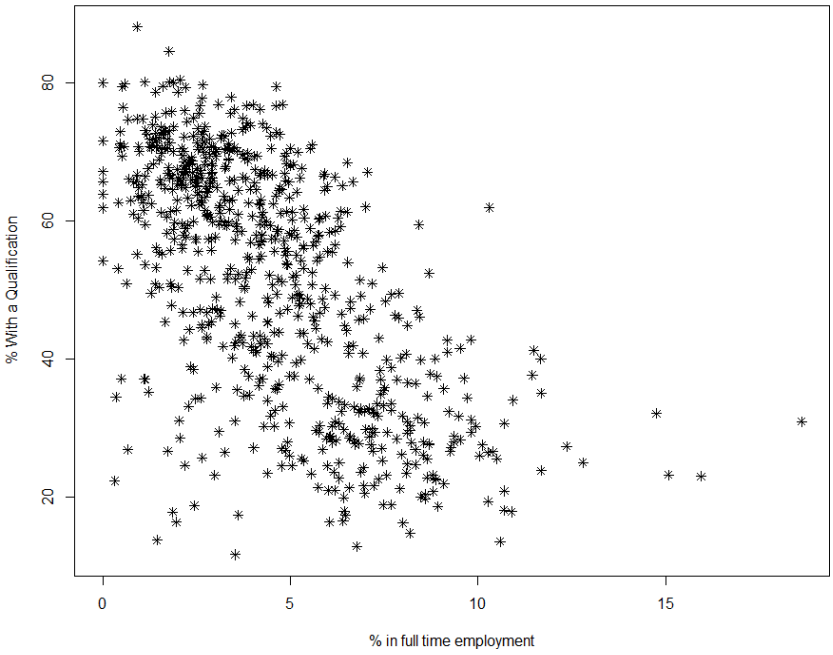


Figure 19. Enhanced simple scatter plot between two variables – employment and qualification, using plot() function with *pch* parameter.

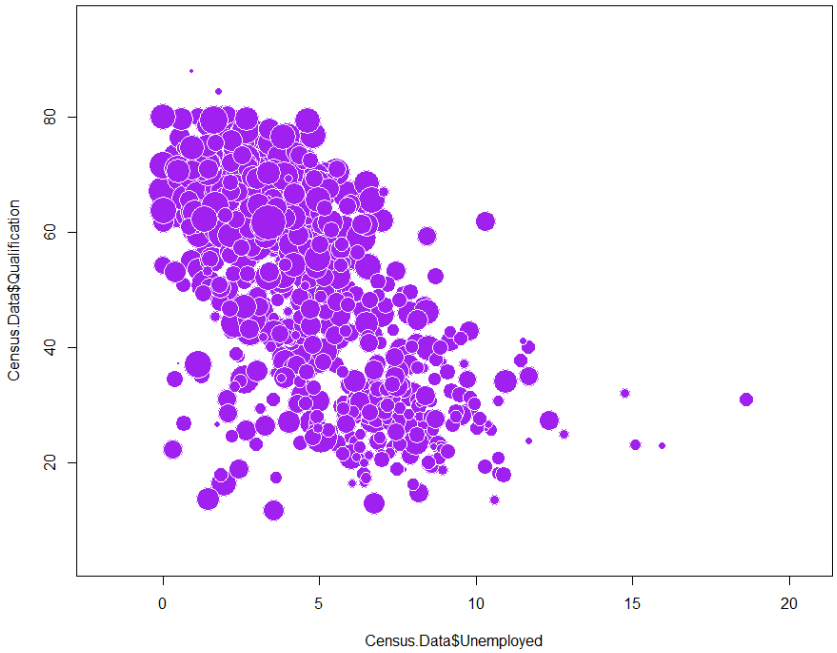


Figure 20. Proportional symbol plot between variables - Unemployed and Qualification, using symbols() function.

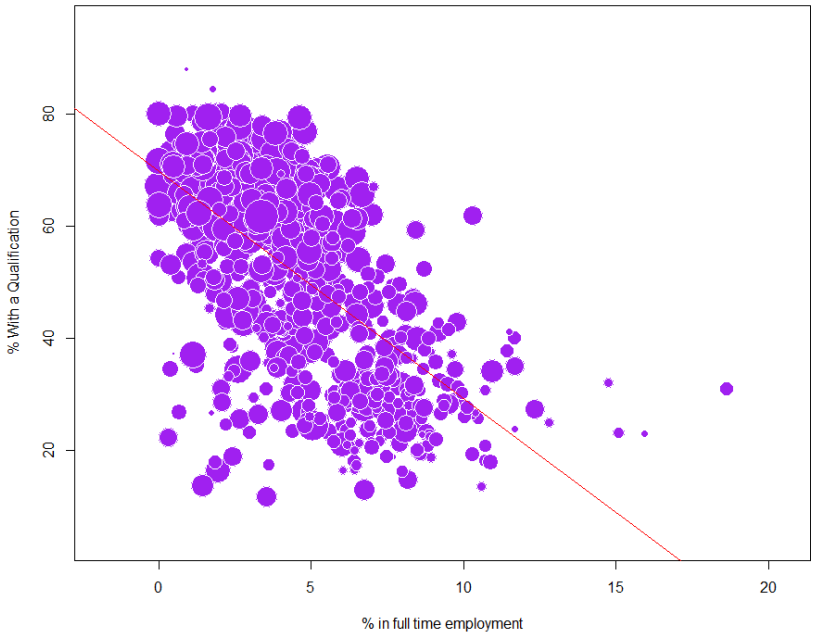


Figure 21. Proportional symbol plot with regression line for relationship between variables - Unemployed and Qualification.

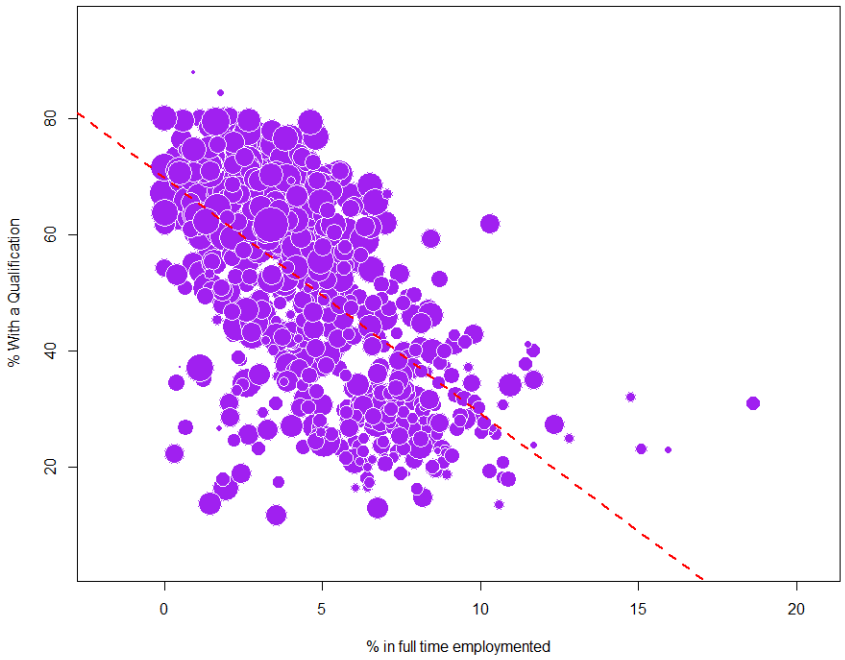


Figure 22. Proportional symbol plot with dotted regression line showing relationship between variables - Unemployed and Qualification.

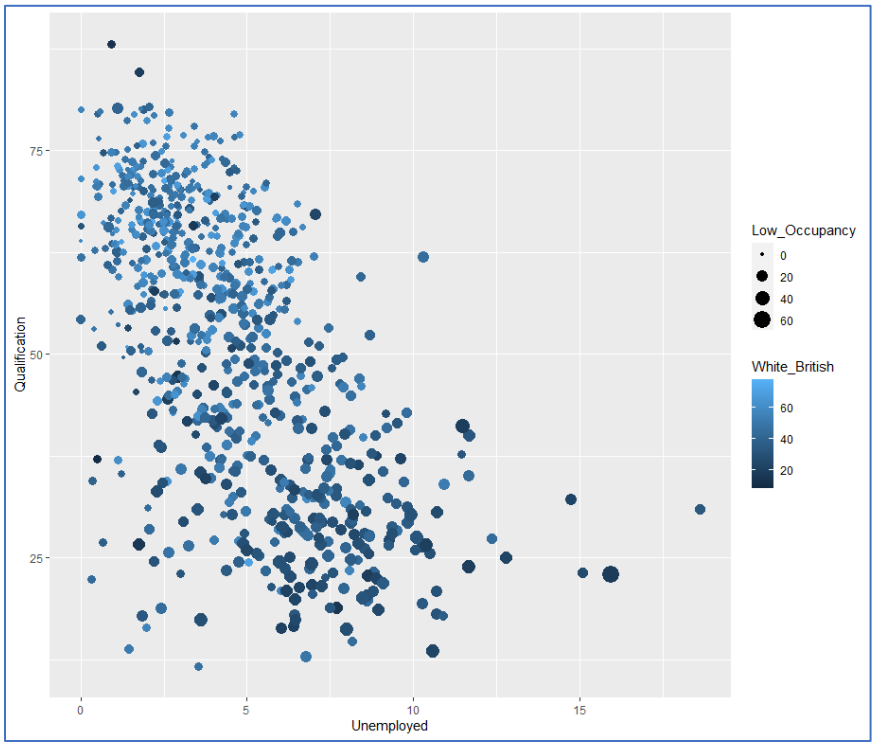


Figure 23. Two-dimensional chart to visualize four different variables, created using ggplot() and geom_point() functions.

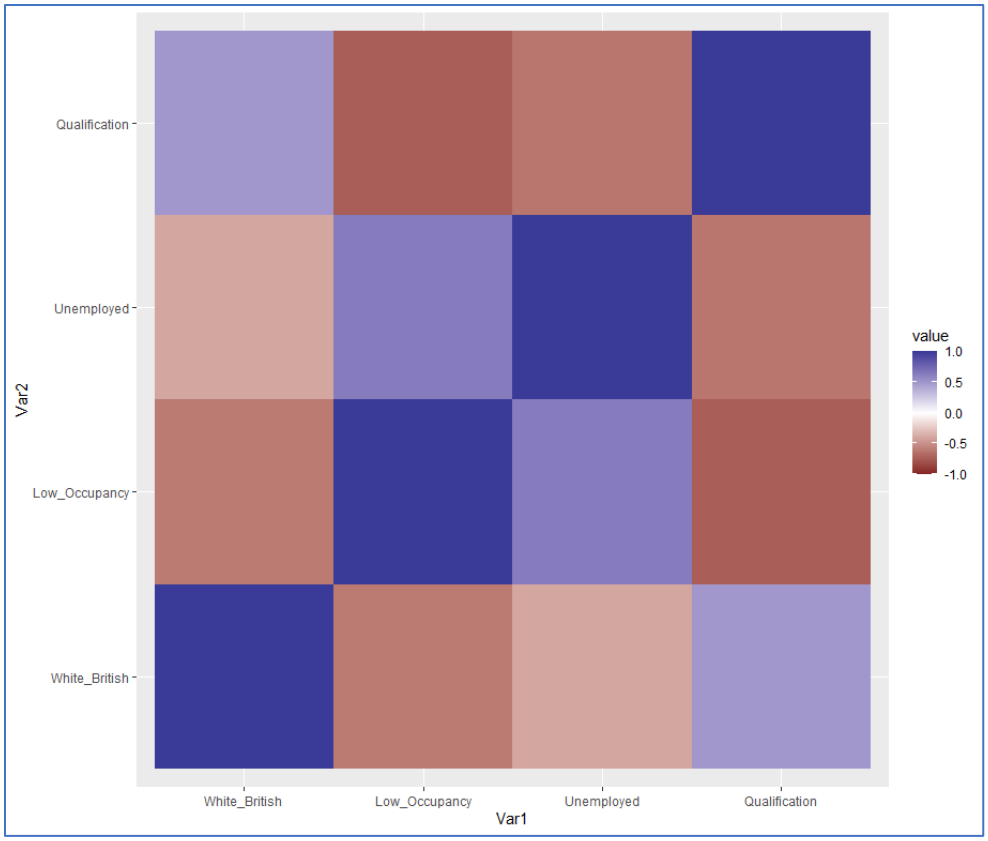


Figure 24. Heat map of correlation matrix between four variables, created using qplot() function.

Figure 25. Camden_oa11 shapefile displayed using plot() function.

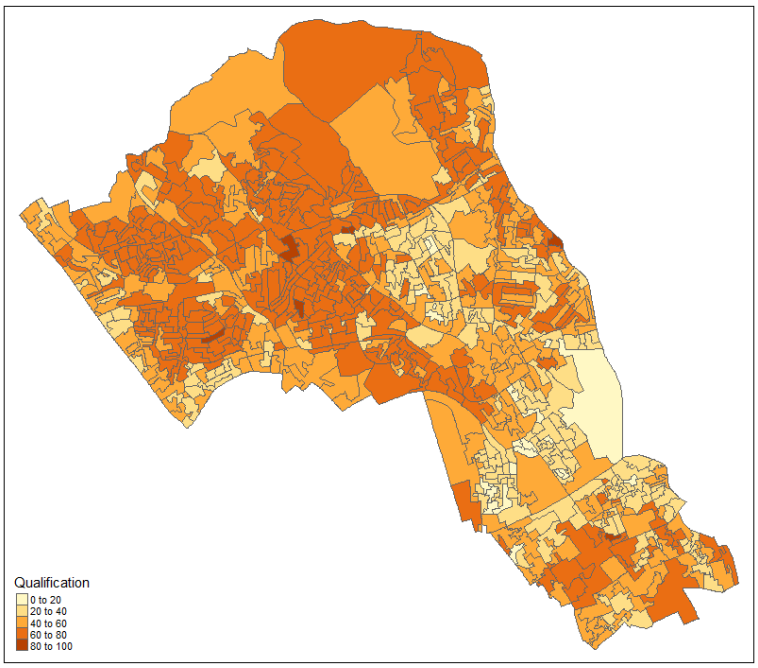


Figure 26. Map showing qualification data created using qtm() function.

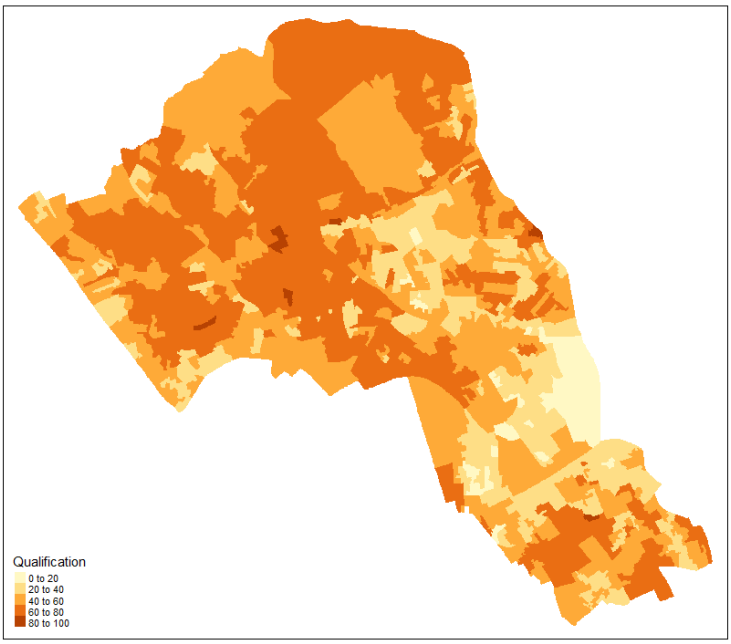


Figure 27. Choropleth map of qualification variable created using tm_shape() and tm_fill() functions.

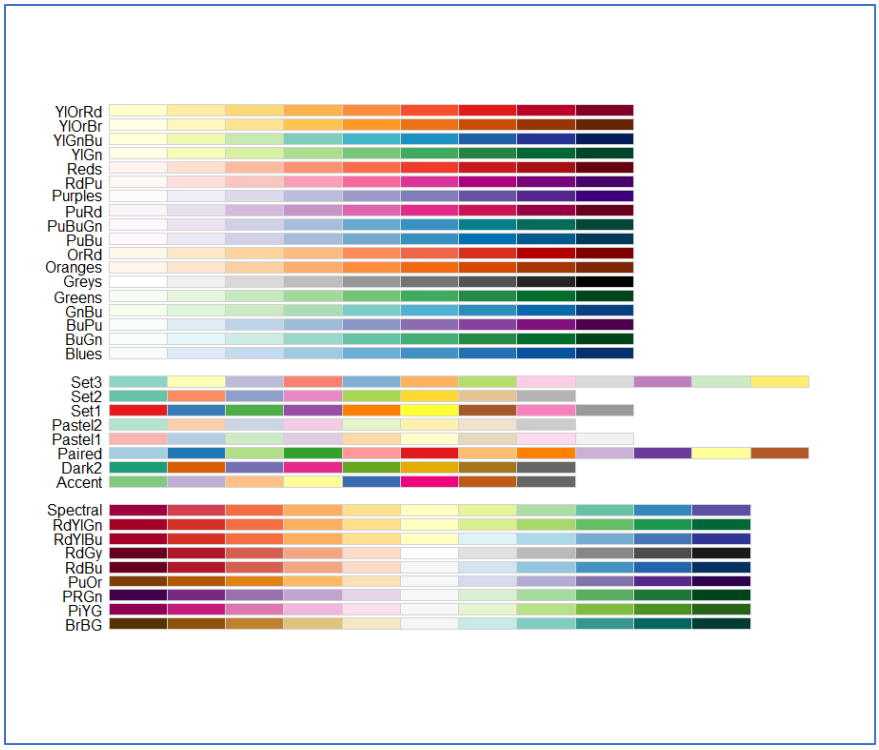


Figure 28. Predefined color ramp displayed using display.brewer.all() function.

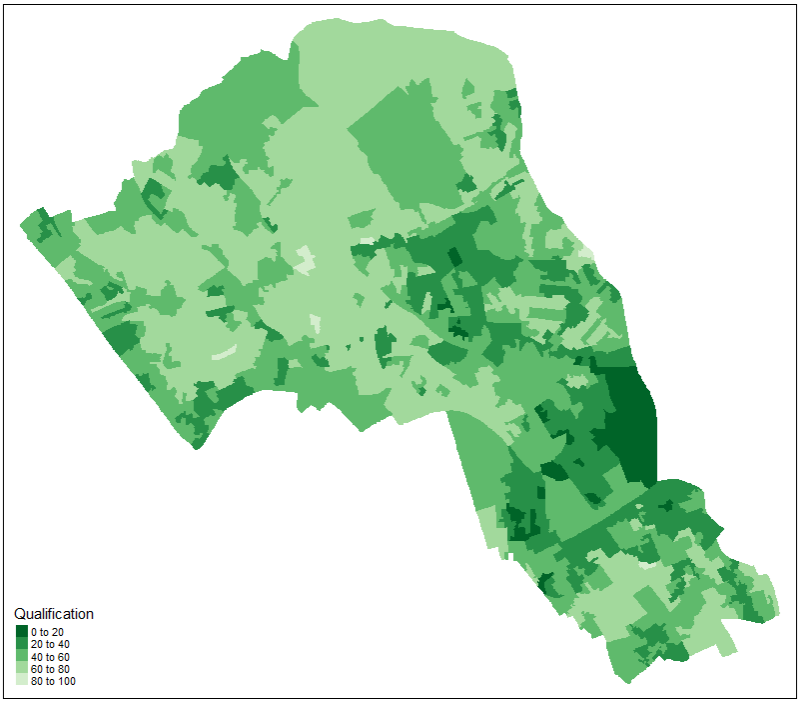


Figure 29. Choropleth map of qualification variable edited using *palette* parameter in tm_fill() function.

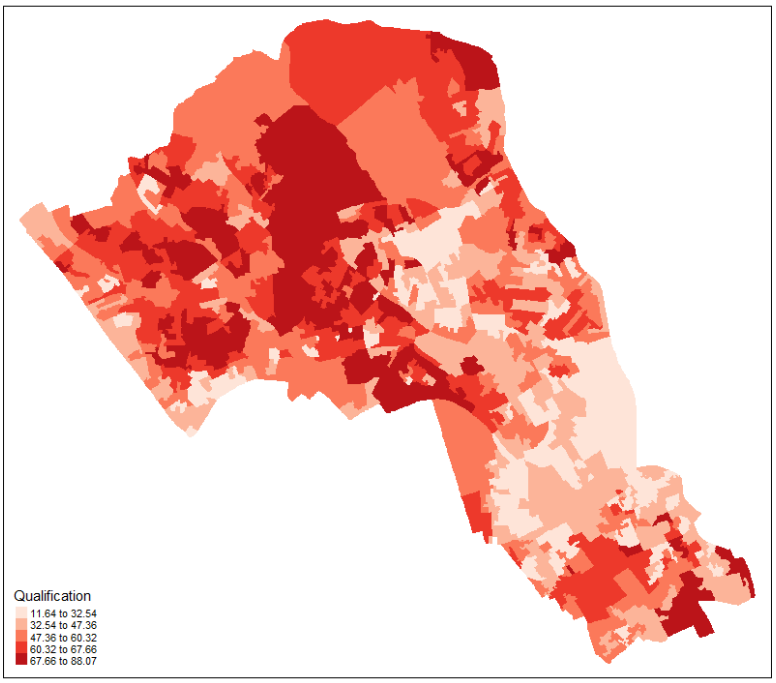


Figure 30. Choropleth map of qualification variable edited using number of intervals *(n)* parameter in tm_fill() function.

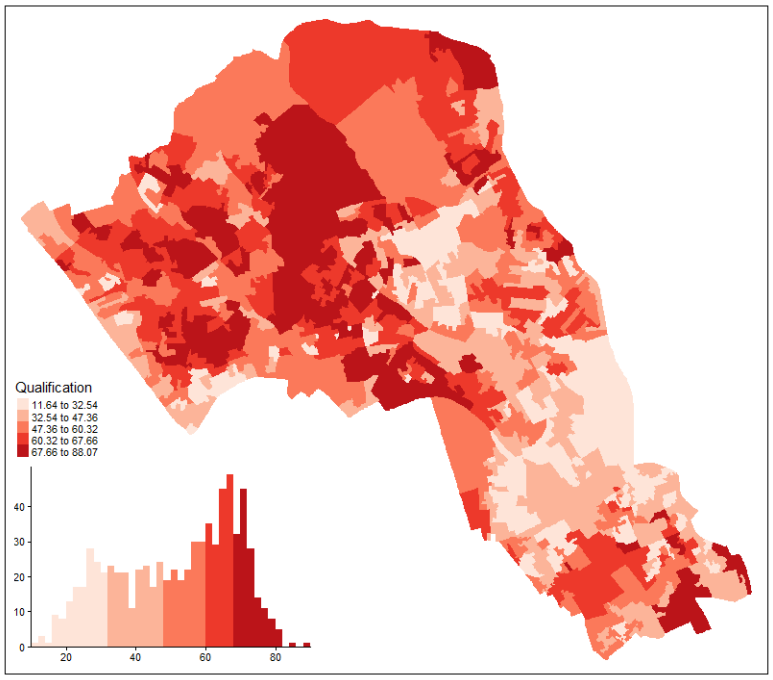


Figure 31. Choropleth map of qualification variable with quantile classification and histogram.

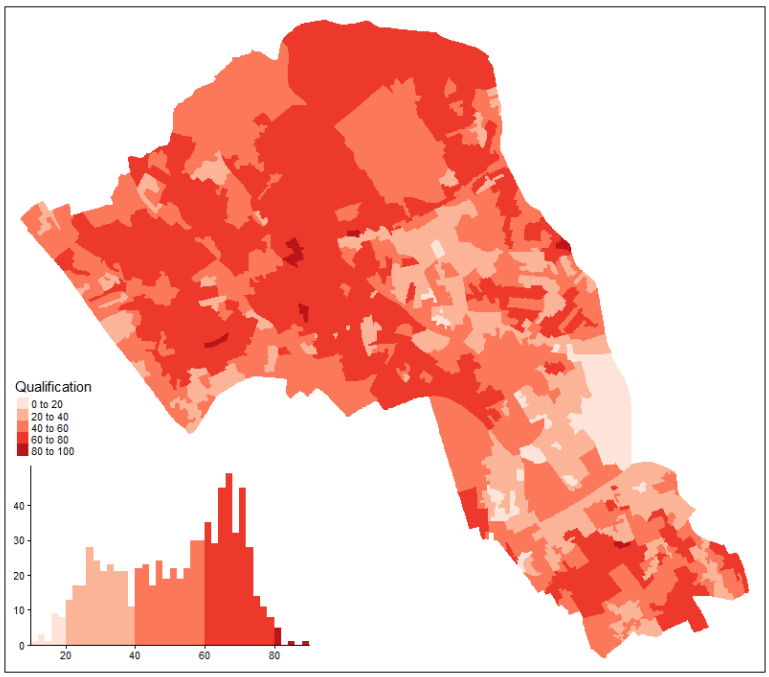


Figure 32. Choropleth map of qualification variable with pretty breaks classification and histogram.

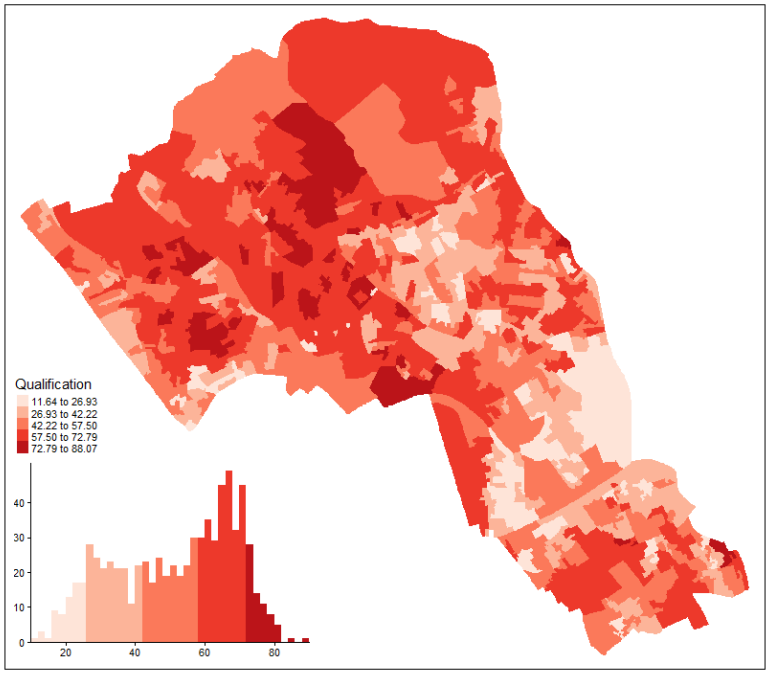


Figure 33. Choropleth map of qualification variable with equal intervals classification and histogram.

Assignment 1: Spatial Analysis and Viz with R (I)

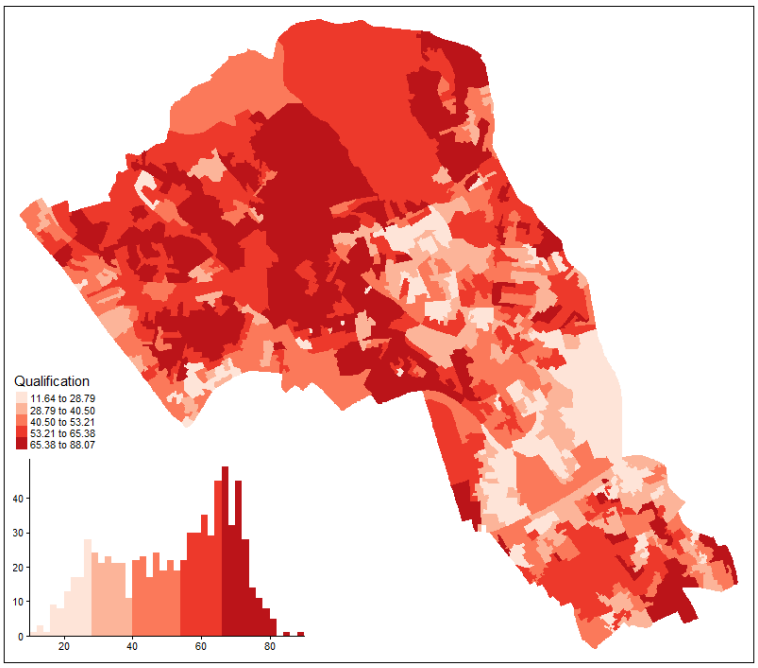


Figure 34. Choropleth map of qualification variable with natural breaks (jenks) classification and histogram.

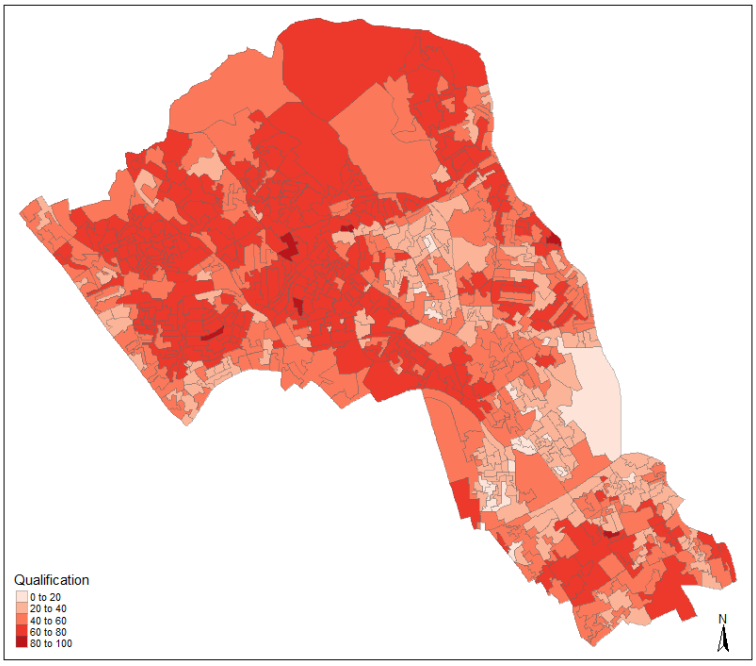


Figure 35. Choropleth map of qualification variable enhanced using tm_borders() and tm_compass() functions.

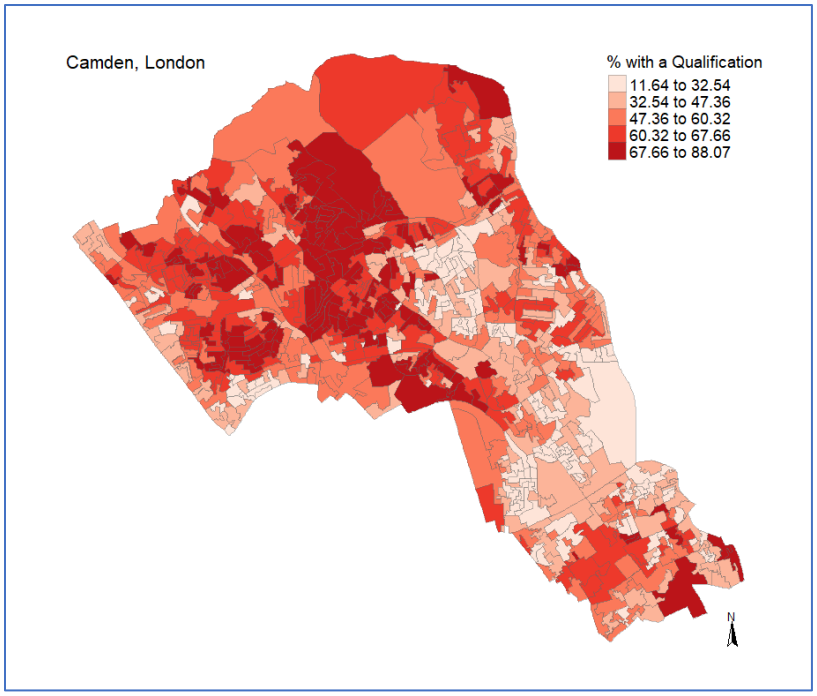


Figure 36. Layout of choropleth map of qualification variable changed using tm_layout() function.

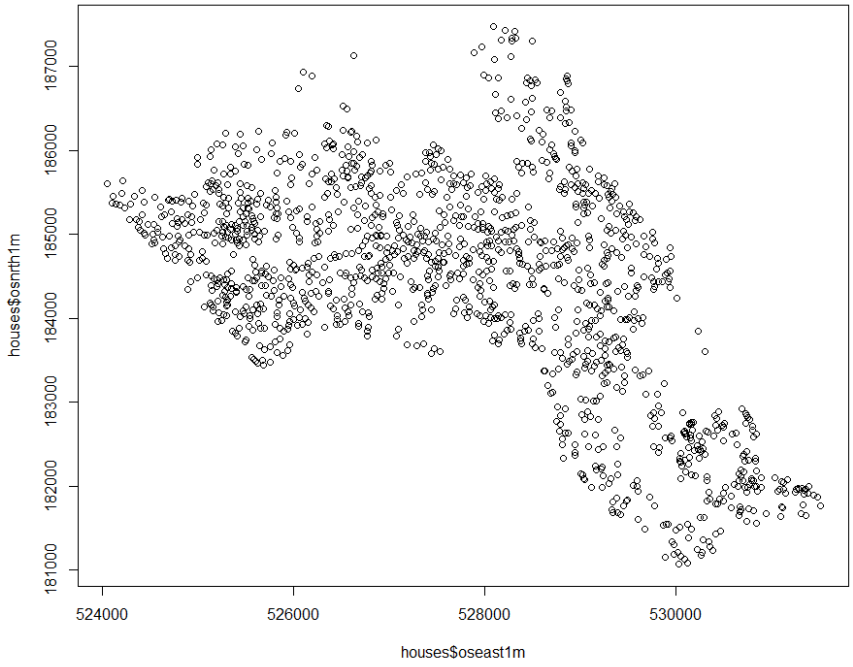


Figure 37. Scatter plot of variables oseast1m and osnrth1m displayed using plot() function.

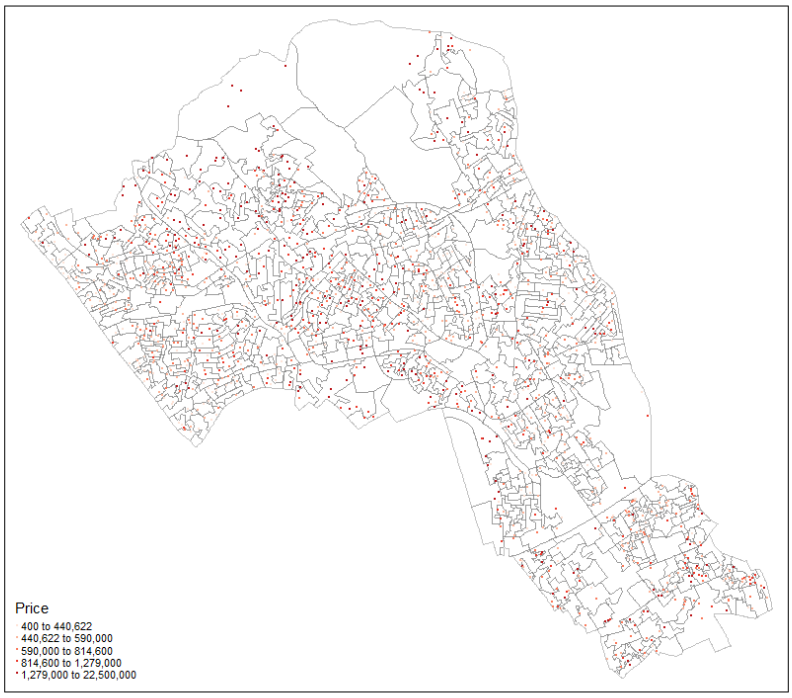


Figure 38. A colored dot map for variable 'price'.

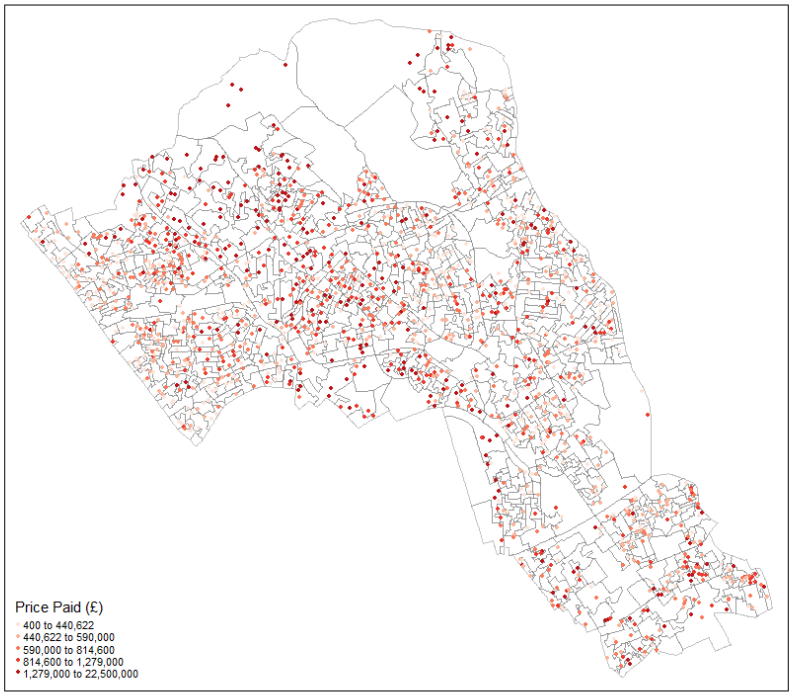


Figure 39. The colored dot map with quantile classification and a legend title created by supplying parameters *style* and *title* to tm_dots() function.

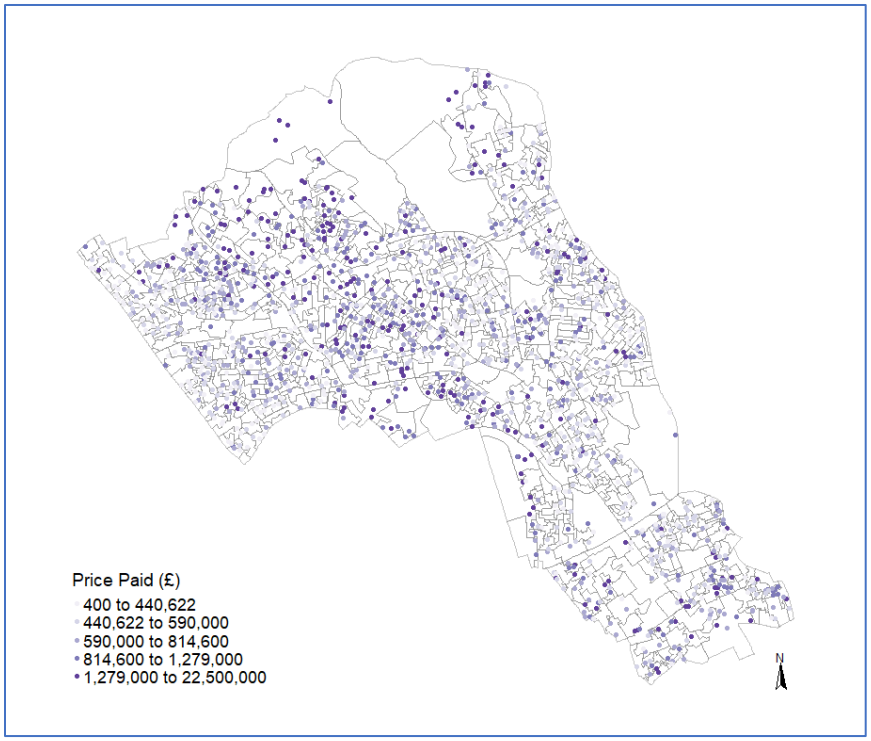


Figure 40. An Enhanced layout of colored dot map for variable 'price' created using tm_compass() and tm_layout() functions.

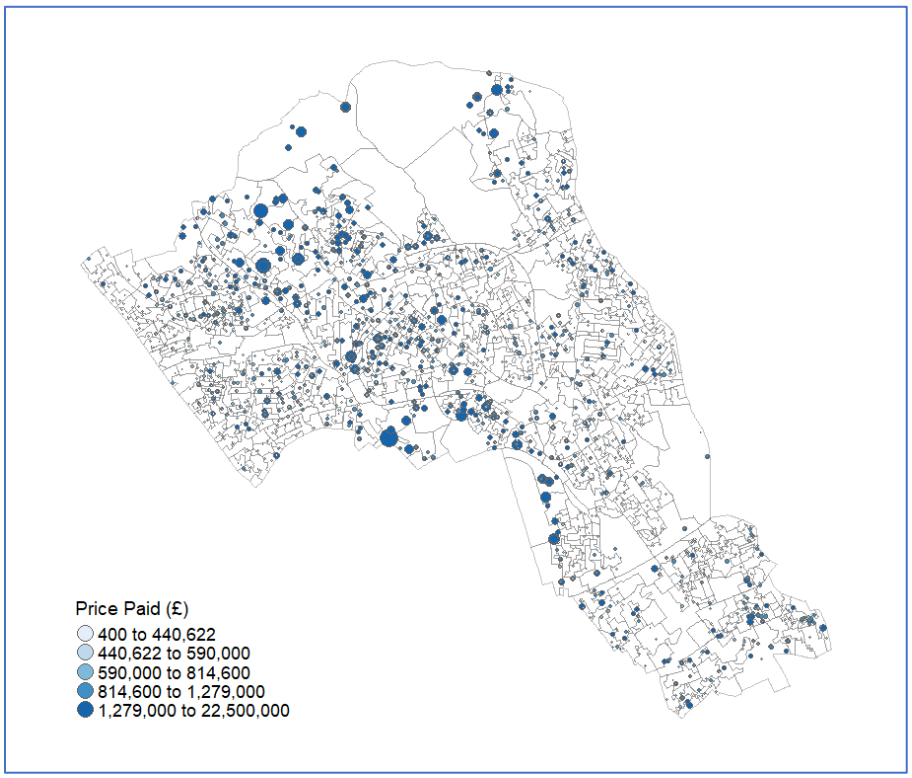


Figure 41. A proportional symbol map for variable 'price' created using tm_bubbles() function.

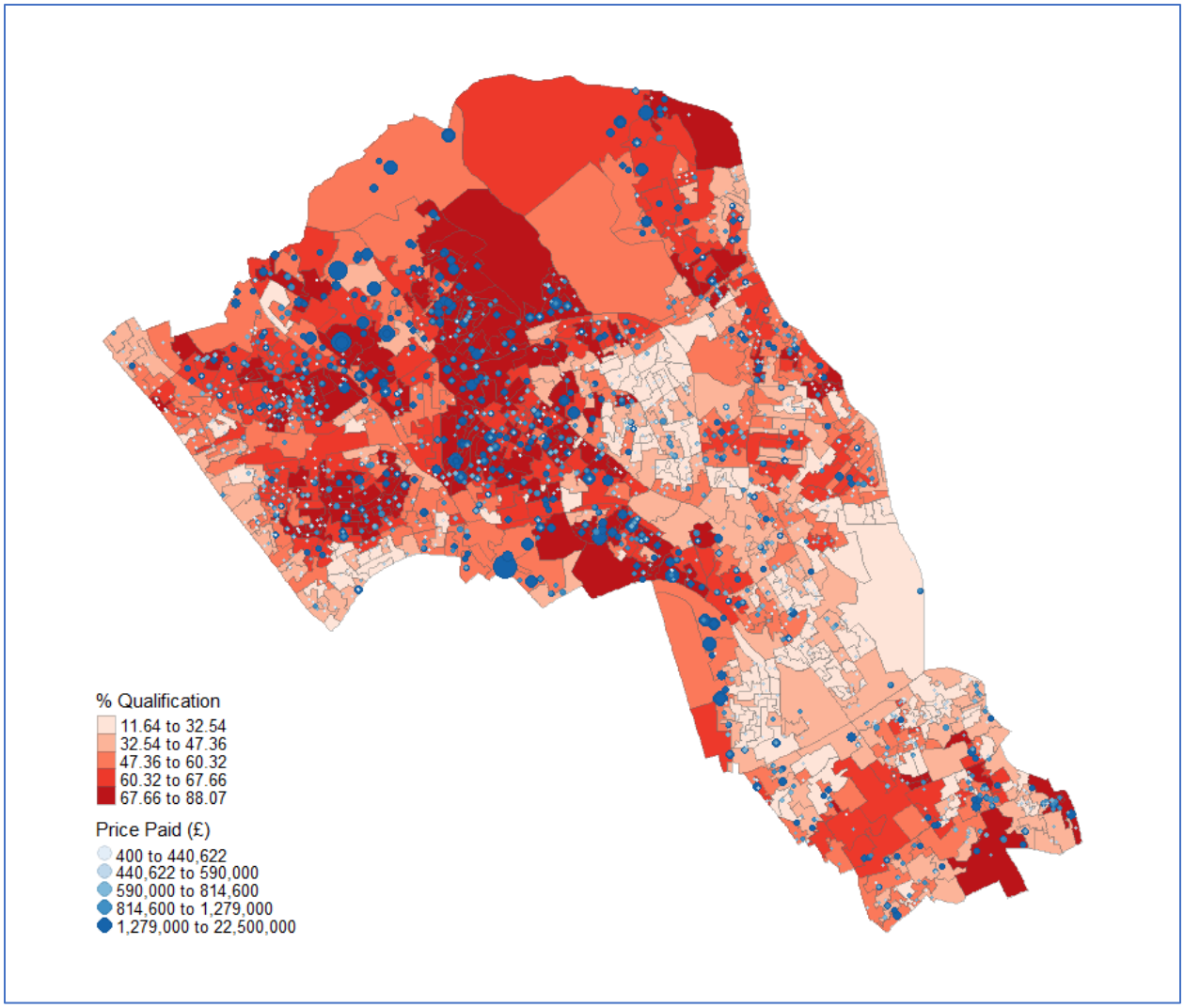


Figure 42. A proportional symbol map for variable 'price' created using tm_bubbles() function displayed over a choropleth map of qualification variable.

3: A 300-500 words reflection summary of what you have learned in this lab.

Practical 1 - An Introduction to R, was a good exercise to refresh my R programming knowledge. I was familiar with loading and viewing data from csv file format. The new and interesting part of this practical was learning how to join the data using merge function. I would say the takeaway from the first exercise was to correctly set the working directory to access the required data.

Practical 2 – Data exploration in R, introduced the descriptive statistics and various techniques to visualize the data. I think it was very straight-forward, simple to understand functions to get descriptive statistics; for instance, using mean() function to get mean of the dataset. This practical also refreshed how to use R for creating histogram and boxplot. I learned a new type of univariate plot in this practical which was violin plot, a combination of histogram and box plot. I also learned the alternate way of installing packages from Tools tab in RStudio.

Data visualization using scatter plot was introduced in Practical 3 – Bivariate Plots in R. plot() function can be used with pch parameter to display a scatter plot with different symbols instead of default hollow circles. In this practical, I learned ggplot2 package can be used to create scatter plot between two variables and with available function aes() we can add additional variables to that two-dimensional chart.

I learned how to calculate the coefficients for linear regression model and functions for obtaining Pearson's correlation coefficient (r) and Spearman's correlation coefficient (rho) in Practical 4 – Finding relationships in R. The R output is very organized and easy to understand. The practical briefly explained the results. For instance, negative value of r or rho indicates there is negative relationship between two variables. The interesting technique I learned in this practical was the heat map of correlation matrix created using qplot() function from the ggplot2 package.

In practical 5, I learned how to load and display spatial data. The libraries that can be used to handle spatial data are rgdal, rgeos, tmap, and leaflet. tm_fill() function can take multiple parameters which helps in enhancing the visualization of spatial data. The main takeway for me was that its important to set a coordinate system of the shapefile. This can be done using functions such as proj4string() and CRS().

Practical 6 – Mapping point data in R, focused on various techniques to map non-spatial data. The non-spatial or point data can be joined to spatial data using merge() function and later tm_shape() function can be used to display the data. Another way is to use SpatialPointsDataFrame() function from the library "sp". It adds spatial attributes to the csv file (non-spatial data). Further, tm_shape() with tm_dots() facilitate mapping specific data as needed. This practical was the most interesting practical for me.

4: Your R codes.

(Lines of code are highlighted in blue)

Practical 1: An Introduction to R

```
> setwd("F:/GEOG 678/Week 1 (19Jan)/camden/tables")

> Ethnicity <- read.csv("KS201EW_oa11.csv")
> Rooms <- read.csv("KS403EW_oa11.csv")
> Qualifications <-read.csv("KS501EW_oa11.csv")
> Employment <-read.csv("KS601EW_oa11.csv")

> View(Employment)

> names(Employment)
 [1] "GeographyCode" "KS601EW0001"   "KS601EW0002"   "KS601EW0003"   "KS601EW0004"   "KS601EW0005"
 [7] "KS601EW0006"   "KS601EW0007"   "KS601EW0008"   "KS601EW0009"   "KS601EW0010"   "KS601EW0011"
[13] "KS601EW0012"   "KS601EW0013"   "KS601EW0014"   "KS601EW0015"   "KS601EW0016"   "KS601EW0017"
[19] "KS601EW0018"   "KS601EW0019"   "KS601EW0020"   "KS601EW0021"   "KS601EW0022"   "KS601EW0023"
[25] "KS601EW0024"   "KS601EW0025"   "KS601EW0026"   "KS601EW0027"   "KS601EW0028"   "KS601EW0029"

> Ethnicity <- Ethnicity[, c(1, 21)]
> Rooms <- Rooms[, c(1, 13)]
> Employment <- Employment[, c(1, 20)]
> Qualifications <- Qualifications[, c(1, 20)]

> names(Employment)[2] <- "Unemployed"
> names(Ethnicity)<- c("OA", "White_British")
> names(Rooms)<- c("OA", "Low_Occupancy")
> names(Employment)<- c("OA", "Unemployed")
> names(Qualifications)<- c("OA", "Qualification")

> merged_data_1 <- merge(Ethnicity, Rooms, by="OA")
> merged_data_2 <- merge(merged_data_1, Employment, by="OA")

> Census.Data <- merge(merged_data_2, Qualifications, by="OA")

> rm(merged_data_1, merged_data_2)

> write.csv(Census.Data, "practical_data.csv", row.names=F)
```

Practical 2: Data exploration in R

```
> Census.Data <-read.csv("practical_data.csv")

> print(Census.Data[1:20,1:5])
          OA White_British Low_Occupancy Unemployed Qualification
1  E00004120      42.35669     6.2937063  1.8939394      73.62637
2  E00004121      47.20000     5.9322034  2.6881720      69.90291
3  E00004122      40.67797     2.9126214  1.2121212      67.58242
4  E00004123      49.66216     0.9259259  2.8037383      60.77586
5  E00004124      51.13636     2.0000000  3.8167939      65.98639
6  E00004125      41.41791     3.9325843  3.8461538      74.20635
7  E00004126      48.54015     5.5555556  4.5454545      62.44726
8  E00004127      48.67925     8.8709677  0.9389671      60.35242
9  E00004128      45.39249     2.4844720  2.1645022      70.07874
10 E00004129      49.05660     3.5211268  4.3103448      66.66667
11 E00004130      38.80597     6.2500000  0.9174312      66.66667
12 E00004131      39.64286     7.5630252  1.8691589      64.47368
13 E00004132      55.88235     4.3478261  3.7974684      73.49398
14 E00004133      41.96078     7.6271186  1.9900498      65.38462
15 E00004134      53.19149     6.0000000  2.7027027      72.89157
16 E00004135      46.85315     4.7619048  3.7313433      74.82014
17 E00004136      59.64912     0.9090909  2.7322404      73.68421
18 E00004137      48.16176     5.4421769  2.7522936      69.06780
19 E00004138      42.22222     2.8169014  4.9723757      58.16327
20 E00004139      17.71772    64.2857143 15.9420290      22.96651

> View(Census.Data)

> head(Census.Data)
         OA White_British Low_Occupancy Unemployed Qualification
1 E00004120      42.35669     6.2937063   1.893939      73.62637
2 E00004121      47.20000     5.9322034   2.688172      69.90291
3 E00004122      40.67797     2.9126214   1.212121      67.58242
4 E00004123      49.66216     0.9259259   2.803738      60.77586
5 E00004124      51.13636     2.0000000   3.816794      65.98639
6 E00004125      41.41791     3.9325843   3.846154      74.20635

> tail(Census.Data)
           OA White_British Low_Occupancy Unemployed Qualification
744 E00174675     37.354086      9.401709   2.714932      52.81385
745 E00174676      7.881773      9.868421   0.500000      37.12871
746 E00174677     22.520107      8.125000   4.528302      50.67568
747 E00174678     23.949580      6.194690   1.421801      53.21101
748 E00174679     24.271845      4.081633   1.663894      45.34884
749 E00174680     36.514523     25.274725   8.108108      24.74227

> ncol(Census.Data)
[1] 5

> nrow(Census.Data)
[1] 749

> names(Census.Data)
[1] "OA"             "White_British" "Low_Occupancy" "Unemployed"    "Qualification"

> mean(Census.Data$Unemployed)
[1] 4.510309
```

```
> median(Census.Data$Unemployed)
[1] 4.186047

> range(Census.Data$Unemployed)
[1]  0.00000 18.62348

> summary(Census.Data)
      OA            White_British    Low_Occupancy      Unemployed     Qualification
 Length:749         Min.   : 7.882   Min.   : 0.000   Min.   : 0.000   Min.   :11.64
 Class :character   1st Qu.:35.915   1st Qu.: 6.015   1st Qu.: 2.500   1st Qu.:36.32
 Mode  :character   Median :44.541   Median :10.000   Median : 4.186   Median :55.10
                    Mean   :44.832   Mean   :11.597   Mean   : 4.510   Mean   :51.43
                    3rd Qu.:54.472   3rd Qu.:16.107   3rd Qu.: 6.158   3rd Qu.:66.23
                    Max.   :78.035   Max.   :64.286   Max.   :18.623   Max.   :88.07

> hist(Census.Data$Unemployed)
> hist(Census.Data$Unemployed, breaks=20, col= "blue", main="% in full-time employment", xlab="Percentage")

> boxplot(Census.Data[,2:5])

> library(vioplot)

> vioplot(Census.Data$Unemployed, Census.Data$Qualification, Census.Data$White_British,
+         Census.Data$Low_Occupancy, ylim=c(0,100),
+         col = "dodgerblue", rectCol="dodgerblue3", colMed="dodgerblue4")


> vioplot(Census.Data$Unemployed, Census.Data$Qualification, Census.Data$White_British,
+         Census.Data$Low_Occupancy, ylim=c(0,100),
+         col = "dodgerblue", rectCol="dodgerblue3", colMed="dodgerblue4",
+         names=c("Unemployed", "Qualifications", "White British", "Occupancy"))


> vioplot(Census.Data$Unemployed, Census.Data$Qualification, Census.Data$White_British,
+         Census.Data$Low_Occupancy, ylim=c(0,100),
+         col = "blue", rectCol="dodgerblue3", colMed="dodgerblue4",
+         names=c("Unemployed", "Qualifications", "White British", "Occupancy"))


> vioplot(Census.Data$Unemployed, Census.Data$Qualification, Census.Data$White_British,
+         Census.Data$Low_Occupancy, ylim=c(0,100),
+         col = "pink", rectCol="dodgerblue3", colMed="dodgerblue4",
+         names=c("Unemployed", "Qualifications", "White British", "Occupancy"))
```

Practical 3: Bivariate Plots in R

```
> plot(Census.Data$Unemployed,Census.Data$Qualification)


> plot(Census.Data$Unemployed,Census.Data$Qualification, xlab="% in full time employment",
+      ylab="% With a Qualification")


> plot(Census.Data$Unemployed,Census.Data$Qualification, xlab="% in full time employment",
+      ylab="% With a Qualification",pch = 8)


> symbols(Census.Data$Unemployed,Census.Data$Qualification,
+         circles = Census.Data$White_British,
+         fg="white", bg ="purple", inches = 0.2)

integer(0)


> symbols(Census.Data$Unemployed, Census.Data$Qualification,
+         circles = Census.Data$White_British,
+         fg="white", bg ="purple", inches = 0.2, xlab="% in full time employment",
+         ylab="% With a Qualification") +
+     # adds a regression line, sets the colour to red
+     abline(lm(Census.Data$Qualification~ Census.Data$Unemployed), col="red")
integer(0)


> symbols(Census.Data$Unemployed, Census.Data$Qualification,
+         circles = Census.Data$White_British,
+         fg="white", bg ="purple", inches = 0.2, xlab="% in full time employmented",
+         ylab="% With a Qualification") +
+     abline(lm(Census.Data$Qualification~ Census.Data$Unemployed), col="red", lwd=2, lty=2)
integer(0)


> library("ggplot2")

> p <- ggplot(Census.Data, aes(Unemployed,Qualification))
> p + geom_point(aes(colour = White_British, size = Low_Occupancy))
```

Practical 4: Finding Relationships in R

```
> cor(Census.Data$Unemployed, Census.Data$Qualification)
[1] -0.624431

> cor.test(Census.Data$Unemployed, Census.Data$Qualification)

        Pearson's product-moment correlation

data:  Census.Data$Unemployed and Census.Data$Qualification
t = -21.85, df = 747, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.6662641 -0.5786800
```

```
sample estimates:
      cor
-0.624431


> cor.test(Census.Data$Unemployed, Census.Data$Qualification, method="spearman")

        Spearman's rank correlation rho

data:  Census.Data$Unemployed and Census.Data$Qualification
S = 113733998, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.6240406

> data1 <- Census.Data[,2:5]


> cor(data1)
              White_British Low_Occupancy Unemployed Qualification
White_British     1.0000000    -0.6006639 -0.3984454     0.4992319
Low_Occupancy    -0.6006639     1.0000000  0.6408021    -0.7347354
Unemployed       -0.3984454     0.6408021  1.0000000    -0.6244310
Qualification     0.4992319    -0.7347354 -0.6244310     1.0000000


> round(cor(data1),2)
              White_British Low_Occupancy Unemployed Qualification
White_British           1.0         -0.60      -0.40          0.50
Low_Occupancy          -0.6          1.00       0.64         -0.73
Unemployed             -0.4          0.64       1.00         -0.62
Qualification           0.5         -0.73      -0.62          1.00


> library(ggplot2)
> library(reshape2)


> qplot(x=Var1, y=Var2, data=melt(cor(data1, use="p")), fill=value, geom="tile") +
+     scale_fill_gradient2(limits=c(-1, 1))


> model_1 <- lm(Census.Data$Qualification~ Census.Data$Unemployed)


> plot(Census.Data$Unemployed, Census.Data$Qualification, xlab="% Unemployed",
+      ylab="% With a Qualification") + abline (model_1)
integer(0)


> summary(model_1)

Call:
lm(formula = Census.Data$Qualification ~ Census.Data$Unemployed)

Residuals:
    Min      1Q  Median      3Q     Max
-50.172  -9.635   2.339   9.512  36.887

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               69.7740     0.9743   71.61   <2e-16 ***
Census.Data$Unemployed    -4.0672     0.1861  -21.85   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.53 on 747 degrees of freedom
Multiple R-squared:  0.3899,   Adjusted R-squared:  0.3891
F-statistic: 477.4 on 1 and 747 DF,  p-value: < 2.2e-16


> predict(model_1, data.frame(Unemployed = c(15)))
       1        2        3        4        5        6        7        8        9       10       11
62.071002 58.840723 64.844073 58.370695 54.250426 54.131014 51.286840 65.955037 60.970577 52.243071 66.042627
      12       13       14       15       16       17       18       19       20       21       22
62.171789 54.329026 61.680105 58.781624 54.597968 58.661489 58.579930 49.550479  4.935035 56.984133 59.731581
      23       24       25       26       27       28       29       30       31       32       33
49.438126 63.352128 55.503204 60.226628 56.593332 63.451950 57.899027 63.443750 62.071002 60.815451 55.986959
      34       35       36       37       38       39       40       41       42       43       44
58.153490 63.703573 69.773975 50.966253 67.476139 50.679281 27.044953 32.275246 56.821205 48.367818 52.528025
      45       46       47       48       49       50       51       52       53       54       55
40.279232 35.054232 28.019760 65.492744 52.708927 62.458922 44.596257 64.805356 62.731257 -5.970889 68.336813
      56       57       58       59       60       61       62       63       64       65       66
51.132780 53.745720 51.587443 38.488053 40.340509 45.419665 59.070897 59.247183 52.623259 57.889388 58.450069
      67       68       69       70       71       72       73       74       75       76       77
67.207938 60.815451 65.400674 60.370114 68.065081 58.476281 58.476281 55.440337 67.873429 42.457163 59.678164
      78       79       80       81       82       83       84       85       86       87       88
60.445604 59.035573 37.304972 28.938936 37.380587 54.393921 26.264716 59.030508 46.309534 41.933229 51.330962
      89       90       91       92       93       94       95       96       97       98       99
62.478603 64.647291 41.265775 52.243071 51.166009 59.163967 67.575505 62.801684 22.297284 62.761614 38.057513
     100      101      102      103      104      105      106      107      108      109      110
63.419022 45.740699 57.220982 49.438126 49.771501 64.951244 62.051501 46.227202 61.639636 47.094589 47.564174
     111      112      113      114      115      116      117      118      119      120      121
54.779801 57.176547 59.973566 43.534170 68.013296 67.007193 69.773975 27.491516 54.954328 61.496153 61.767735

########## complete result is not copied here #################
```

```
> summary(model_1)

Call:
lm(formula = Census.Data$Qualification ~ Census.Data$Unemployed)

Residuals:
    Min      1Q  Median      3Q     Max
-50.172  -9.635   2.339   9.512  36.887

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)             69.7740     0.9743   71.61   <2e-16 ***
Census.Data$Unemployed  -4.0672     0.1861  -21.85   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.53 on 747 degrees of freedom
Multiple R-squared:  0.3899,   Adjusted R-squared:  0.3891
F-statistic: 477.4 on 1 and 747 DF,  p-value: < 2.2e-16


> confint(model_1, level= 0.95)
                            2.5 %    97.5 %
(Intercept)             67.861262 71.686689
Census.Data$Unemployed -4.432593 -3.701747
```

```
> model_2 <- lm(Census.Data$Qualification~ Census.Data$Unemployed +
+                   Census.Data$White_British)


> summary(model_2)

Call:
lm(formula = Census.Data$Qualification ~ Census.Data$Unemployed +
    Census.Data$White_British)

Residuals:
    Min      1Q  Median      3Q     Max
-50.311  -8.014   1.006   8.958  38.046

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               47.86697    2.33574   20.49   <2e-16 ***
Census.Data$Unemployed    -3.29459    0.19027  -17.32   <2e-16 ***
Census.Data$White_British  0.41092    0.04032   10.19   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.69 on 746 degrees of freedom
Multiple R-squared:  0.4645,   Adjusted R-squared:  0.463
F-statistic: 323.5 on 2 and 746 DF,  p-value: < 2.2e-16
```

Practical 5: Making maps in R

```
> library("rgdal")
> library("rgeos")


> getwd()
[1] "F:/GEOG 678/Week 1 (19Jan)/camden/tables"
```

```
> setwd("F:/GEOG 678/Week 1 (19Jan)/camden/shapefiles")

> Output.Areas<- readOGR(".", "Camden_oa11")

OGR data source with driver: ESRI Shapefile
Source: "F:\GEOG 678\Week 1 (19Jan)\camden\shapefiles", layer: "Camden_oa11"
with 749 features
It has 1 fields

> plot(Output.Areas)

> OA.Census <- merge(Output.Areas, Census.Data, by.x="OA11CD", by.y="OA")

> OA.Census
class       : SpatialPolygonsDataFrame
features    : 749
extent      : 523954.5, 531554.9, 180959.8, 187603.6  (xmin, xmax, ymin, ymax)
crs         : +proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717 +x_0=400000 +y_0=-100000 +ellps=airy +units=m +no_defs
variables   : 5
names       :    OA11CD,    White_British,    Low_Occupancy,       Unemployed,    Qualification
min values  : E00004120, 7.88177339901478,                0,                0, 11.6438356164384
max values  : E00174680, 78.0346820809249, 64.2857142857143, 18.6234817813765, 88.0733944954129


> proj4string(OA.Census) <- CRS("+init=EPSG:27700")

> library(tmap)
> library(leaflet)

> qtm(OA.Census, fill = "Qualification")
```

```
> tm_shape(OA.Census) + tm_fill("Qualification")

> library(RColorBrewer)

> display.brewer.all()

> tm_shape(OA.Census) + tm_fill("Qualification", palette = "-Greens")

> tm_shape(OA.Census) + tm_fill("Qualification", style = "quantile", palette = "Reds")

> tm_shape(OA.Census) + tm_fill("Qualification", style = "quantile", n = 7,
+                               palette = "Reds")

> tm_shape(OA.Census) + tm_fill("Qualification", style = "quantile", n = 5,
+                               palette = "Reds", legend.hist = TRUE)

> tm_shape(OA.Census) + tm_fill("Qualification", style = "pretty", n = 5,
+                               palette = "Reds", legend.hist = TRUE)

> tm_shape(OA.Census) + tm_fill("Qualification", style = "equal", n = 5,
+                               palette = "Reds", legend.hist = TRUE)

> tm_shape(OA.Census) + tm_fill("Qualification", style = "jenks", n = 5,
+                               palette = "Reds", legend.hist = TRUE)

> tm_shape(OA.Census) + tm_fill("Qualification", palette = "Reds") +
+     tm_borders(alpha=.4)

> tm_shape(OA.Census) + tm_fill("Qualification", palette = "Reds") +
+     tm_borders(alpha=.4) +
+     tm_compass()
```

```
> tm_shape(OA.Census) + tm_fill("Qualification", palette = "Reds",
+                               style = "quantile", title = "% with a Qualification") +
+     tm_borders(alpha=.4) +
+     tm_compass() +
+     tm_layout(title = "Camden, London", legend.text.size = 1.1,
+               legend.title.size = 1.4, legend.position = c("right", "top"), frame = FALSE)



> writeOGR(OA.Census, dsn = "C:/Users/Guy/Documents/Teaching/CDRC/Practicals",
+          layer = "Census_OA_Shapefile", driver="ESRI Shapefile")

> writeOGR(OA.Census, dsn = getwd(),
+          layer = "Census_OA_Shapefile", driver="ESRI Shapefile")



> output = readOGR(".", "Census_OA_Shapefile")
OGR data source with driver: ESRI Shapefile
Source: "F:\GEOG 678\Week 1 (19Jan)\camden\shapefiles", layer: "Census_OA_Shapefile"
with 749 features
It has 5 fields



> head(output)
class       : SpatialPolygonsDataFrame
features    : 6
extent      : 524326, 530660.2, 181181.1, 185111.2  (xmin, xmax, ymin, ymax)
crs         : +proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717 +x_0=400000 +y_0=-100000 +ellps=airy +units=m +no_defs
variables   : 5
names       :    OA11CD,          Wht_Brt,          Lw_Occp,          Unmplyd,          Qulfctn
min values  : E00004200, 31.8681318681319, 8.54700854700855, 2.11640211640212, 31.7460317460317
max values  : E00004527, 56.4516129032258, 19.6850393700787, 7.98319327731092, 67.8571428571429
```

Practical 6: Mapping Point Data in R

```
> setwd("F:/GEOG 678/Week 1 (19Jan)/camden/tables")

> Census.Data <-read.csv("practical_data.csv")

> library("rgdal")
> library("rgeos")

> setwd("F:/GEOG 678/Week 1 (19Jan)/camden/shapefiles")

> Output.Areas <- readOGR(".", "Camden_oa11")
OGR data source with driver: ESRI Shapefile
Source: "F:\GEOG 678\Week 1 (19Jan)\camden\shapefiles", layer: "Camden_oa11"
with 749 features
It has 1 fields


> OA.Census
class       : SpatialPolygonsDataFrame
features    : 749
extent      : 523954.5, 531554.9, 180959.8, 187603.6  (xmin, xmax, ymin, ymax)
crs         : +proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717 +x_0=400000 +y_0=-100000 +ellps=airy +units=m +no_defs
variables   : 5
names       :    OA11CD,    White_British,    Low_Occupancy,       Unemployed,    Qualification
min values  : E00004120, 7.88177339901478,                0,                0, 11.6438356164384
max values  : E00174680, 78.0346820809249, 64.2857142857143, 18.6234817813765, 88.0733944954129


> setwd("F:/GEOG 678/Week 1 (19Jan)/camden/tables")

> houses <- read.csv("CamdenHouseSales15.csv")
> houses <- houses[,c(1,2,8,9)]
```

```
> plot(houses$oseast1m, houses$osnrth1m)

> library("sp")

> House.Points <-SpatialPointsDataFrame(houses[,3:4], houses,
+                                       proj4string = CRS("+init=EPSG:27700"))


> library("tmap")

> tm_shape(OA.Census) + tm_borders(alpha=.4)

> tm_shape(OA.Census) + tm_borders(alpha=.4) +
+     tm_shape(House.Points) + tm_dots(col = "Price", palette = "Reds", style = "quantile")

> tm_shape(OA.Census) + tm_borders(alpha=.4) +
+     tm_shape(House.Points) + tm_dots(col = "Price", scale = 1.5, palette = "Reds",
+                                      style = "quantile", title = "Price Paid (£)")

> tm_shape(OA.Census) + tm_borders(alpha=.4) +
+     tm_shape(House.Points) + tm_dots(col = "Price", scale = 2, palette = "Reds",
+                                      style = "quantile", title = "Price Paid (£)")

> tm_shape(OA.Census) + tm_borders(alpha=.4) +
+     tm_shape(House.Points) + tm_dots(col = "Price", scale = 2.5, palette = "Purples",
+                                      style = "quantile", title = "Price Paid (£)") +
+     tm_compass() +
+     tm_layout(legend.text.size = 1.1, legend.title.size = 1.4, frame = FALSE)

> tm_shape(OA.Census) + tm_borders(alpha=.4) +
+     tm_shape(House.Points) + tm_bubbles(size = "Price", col = "Price",
+                                         palette = "Blues", style = "quantile",
+                                         legend.size.show = FALSE,
+                                         title.col = "Price Paid (£)") +
+     tm_layout(legend.text.size = 1.1, legend.title.size = 1.4, frame = FALSE)


> tm_shape(OA.Census) + tm_fill("Qualification", palette = "Reds",
+                               style = "quantile", title = "% Qualification") +
+     tm_borders(alpha=.4) +
+     tm_shape(House.Points) + tm_bubbles(size = "Price", col = "Price",
+                                         palette = "Blues", style = "quantile",
+                                         legend.size.show = FALSE,
+                                         title.col = "Price Paid (£)",
+                                         border.col = "black", border.lwd = 0.1,
+                                         border.alpha = 0.1) +
+     tm_layout(legend.text.size = 0.8, legend.title.size = 1.1, frame = FALSE)


> writeOGR(House.Points, dsn = getwd(),
+          layer = "Camden_house_sales", driver="ESRI Shapefile")
```