

## Lab 1

### Environmental Analysis in R

*Srishti Roy Chowdhury*

---

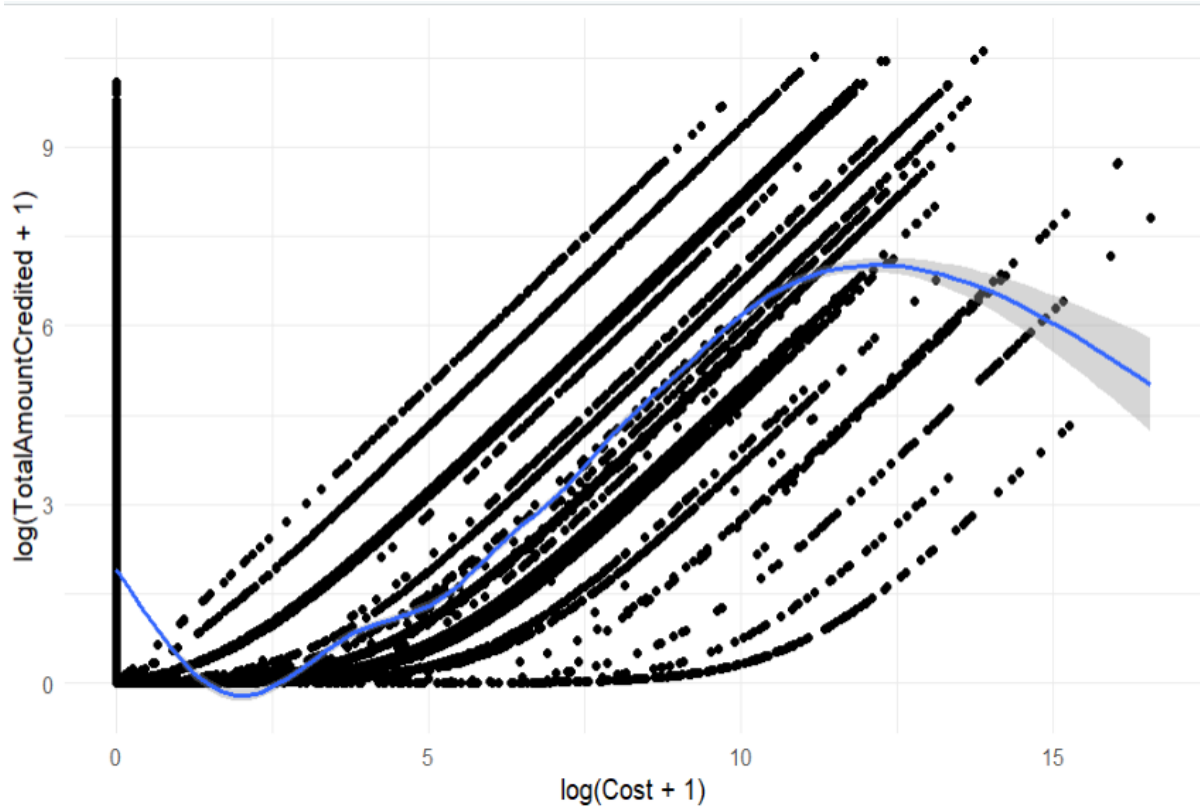
#### Task- 1.1

```
> bmps %>% group_by(StateAbbreviation) %>% summarize(totalCost = sum(Cost,na.rm = TRUE))  
# A tibble: 7 × 2  
  StateAbbreviation totalCost  
    <chr>          <dbl>  
1 DC              15623736.  
2 DE              41329130.  
3 MD              627164538.  
4 NY              15273818.  
5 PA              562922554.  
6 VA              723866417.  
7 WV              83023958.
```

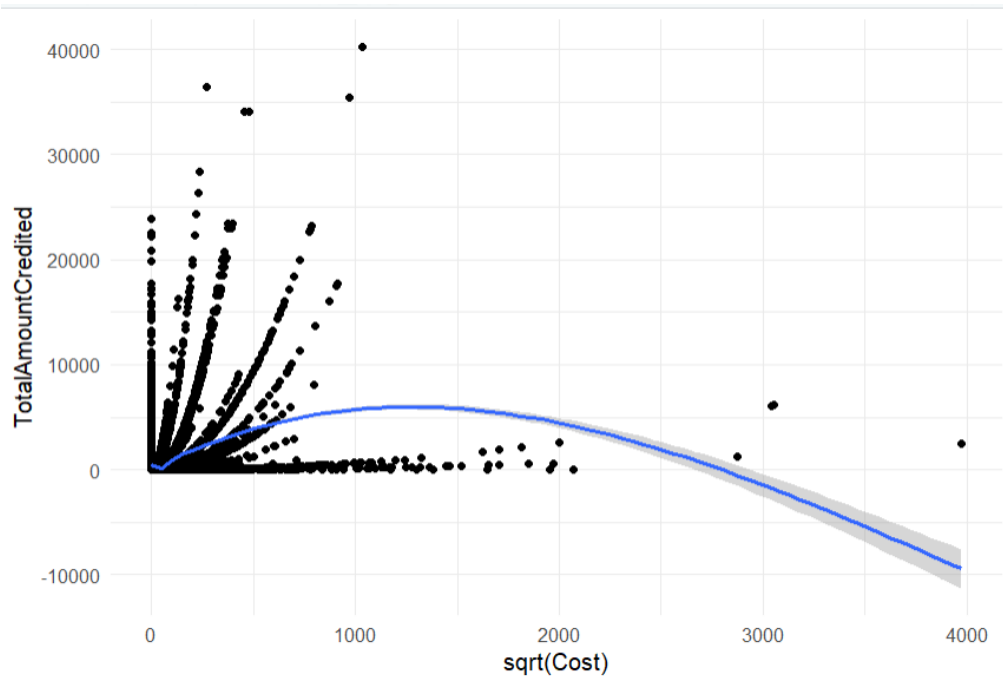
#### Task- 1.2

I tried multiple data transformations and got confused between the two, which represent a better normal distribution.

```
> bmps %>% dplyr::filter(, Unit=="Acres") %>%  
+   ggplot(., aes(x = log(Cost+1), y = log(TotalAmountCredited+1))) +  
+   geom_point() +  
+   geom_smooth() +  
+   theme_minimal()
```

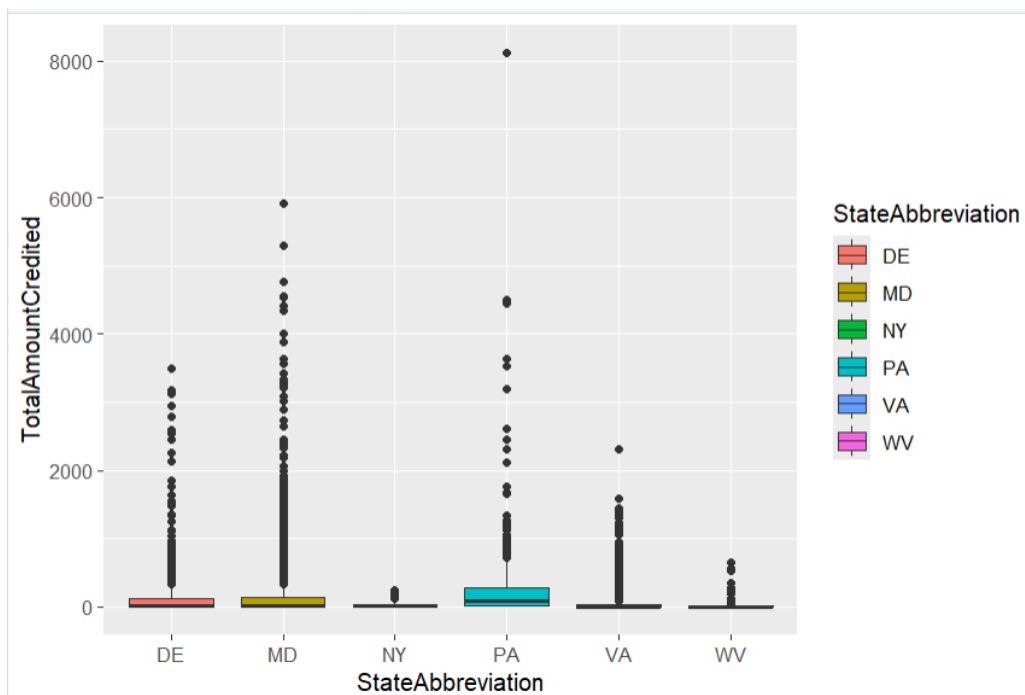


```
> bmps %>% dplyr::filter(., Unit=="Acres") %>%
+   ggplot(., aes(x = sqrt(Cost), y = TotalAmountCredited)) +
+   geom_point() +
+   geom_smooth() +
+   theme_minimal()
```



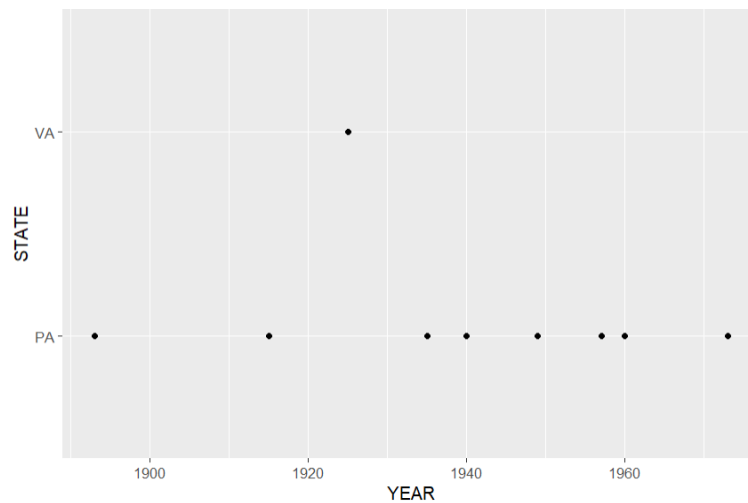
### Task- 1.3

```
> bmps13 <- bmps %>% mutate(., trim_bmp=stringr::str_sub(BMPShortName, 1, 9))
> bmps13 %>% filter(trim_bmp=="covercrop") %>%
+   ggplot(., aes(x = StateAbbreviation, y = TotalAmountCredited)) +
+   geom_boxplot(aes(fill = StateAbbreviation))
```



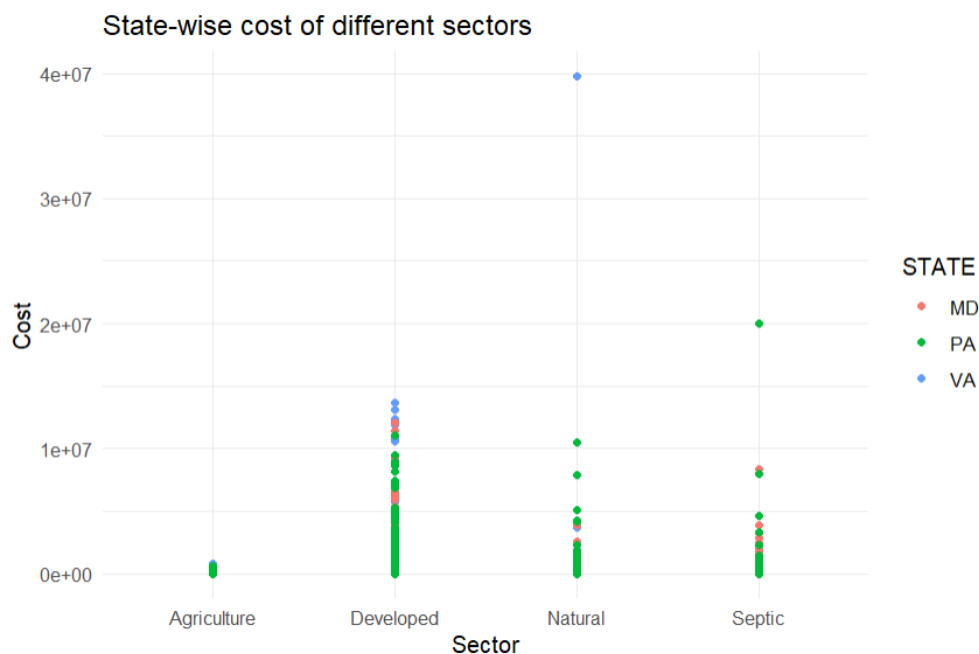
### Task- 1.4

```
> dams %>% filter(YEAR!=0) %>%  
+   ggplot(., aes(x = YEAR, y = STATE)) +  
+   geom_point()
```



### Task- 1.5

```
> joined_table <- left_join(dams, bmps, by = c("STATE" = "StateAbbreviation"))  
Warning message:  
In sf_column %in% names(g) :  
  Detected an unexpected many-to-many relationship between `x` and `y`.  
i Row 1 of `x` matches multiple rows in `y`.  
i Row 18856 of `y` matches multiple rows in `x`.  
i If a many-to-many relationship is expected, set `relationship = "many-to-many"` to  
  silence this warning.  
> joined_table %>% ggplot(., aes(x = Sector, y = Cost, color=STATE)) +  
+   geom_point() +  
+   theme_minimal() +  
+   labs(title = "State-wise cost of different sectors",  
+         x = "Sector",  
+         y = "Cost")  
Warning message:  
Removed 135567 rows containing missing values or values outside the scale range  
(`geom_point()`).
```



## Task- 2.1

```
> streams %>% dplyr::arrange(desc(LengthKM)) %>% head(5)
Simple feature collection with 5 features and 89 fields
Geometry type: LINESTRING
Dimension:      XYZ
Bounding box:   xmin: -79.06165 ymin: 37.96144 xmax: -76.49969 ymax: 40.43248
z_range:        zmin: 0 zmax: 0
Geodetic CRS:   WGS 84
# A tibble: 5 × 90
  OBJECTID_1  ComID Permanent_ FDate      Resolution GNIS_ID GNIS_Name      LengthKM
  <int>      <int> <chr>      <date>      <chr>      <chr> <chr>      <dbl>
1     7470 57464201 NA      1970-01-01 NA      NA      Little Chiques ... 6.85
2     9032 56409913 NA      1970-01-01 NA      NA      Toms Run          6.54
3     6693 41377791 NA      1970-01-01 NA      NA      Johns Run         5.43
4     7794 65839879 NA      1970-01-01 NA      NA      Little Trough C... 5.07
5     7848 65842019 NA      1970-01-01 NA      NA      Great Trough Cr... 4.89
```

## Task- 2.2

```
> counties <- counties %>% sf::st_make_valid()
> counties %>% sf::st_crs() == streams %>% sf::st_crs()
[1] TRUE
>
> streams.counties <- sf::st_join(streams, counties)
st_as_s2(): dropping Z and/or M coordinate
> streams.countie_sum <- streams.counties %>% group_by(COUNTYFP10) %>% summarise(TotalLength=
sum(LengthKM))
st_as_s2(): dropping Z and/or M coordinate
> streams.countie_sum %>% dplyr::arrange(desc(TotalLength)) %>% head(3)
Simple feature collection with 3 features and 2 fields
Geometry type: MULTILINESTRING
Dimension:      XY
Bounding box:   xmin: -79.23801 ymin: 37.88782 xmax: -77.96634 ymax: 41.62481
Geodetic CRS:   WGS 84
# A tibble: 3 × 3
  COUNTYFP10 TotalLength geometry
  <chr>      <dbl>      <MULTILINESTRING [°]>
1 015        425. ((-79.20884 37.98276, -79.20882 37.98262, -79.20881 37.98249, ...
2 061        309. ((-78.01929 40.36854, -78.01908 40.36834, -78.01875 40.36801, ...
3 023        242. ((-78.40534 41.57611, -78.40511 41.57597, -78.40495 41.57594, ...
```

### Task- 2.3

Every time I tried it was taking a lot of time to produce the map, and sometimes, R stopped working and so did my computer which made me unable to attach the map to this file.

```

Console Terminal × Background Jobs ×
R • R 4.4.2 · ~/1-2/R/R_labs/Lab2/ ↗
> bmps <- bmps %>% mutate(., FIPS.trimmed = stringr::str_sub(GeographyName, 1, 5))
> counties.bmps <- left_join(counties, bmps, by = c("GEOID10" = "FIPS.trimmed"))
> glimpse(counties.bmps)
Rows: 69,614
Columns: 39
$ OBJECTID          <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
$ STATEFP10         <chr> "51", "51", "51", "51", "51", "51", "51", "51", "51", "51", "51", "51",...
$ COUNTYFP10        <chr> "540", "540", "540", "540", "540", "540", "540", "540", "540", "540", "5...
$ COUNTYNYS10       <chr> "01789068", "01789068", "01789068", "01789068", "01789068", "01789068", "0178...
$ GEOID10            <chr> "51540", "51540", "51540", "51540", "51540", "51540", "51540", "51540", "5...
$ NAME10             <chr> "Charlottesville", "Charlottesville", "Charlottesville", "Charlot...
$ NAMEDSAD10        <chr> "Charlottesville city", "Charlottesville city", "Charlottesville ...
$ LSAD10             <chr> "25", "25", "25", "25", "25", "25", "25", "25", "25", "25", "25", "25"...
$ CLASSFP10         <chr> "C7", "C7", "C7", "C7", "C7", "C7", "C7", "C7", "C7", "C7", "C7", "C7"...
$ MTFCC10           <chr> "G4020", "G4020", "G4020", "G4020", "G4020", "G4020", "G4020", "G...
$ CSAFP10            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ CBSAFP10           <chr> "16820", "16820", "16820", "16820", "16820", "16820", "16820", "1...
$ METDIVFP10        <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ FUNCSTAT10        <chr> "F", "F", "F", "F", "F", "F", "F", "F", "F", "F", "F", "F", "F", "...
$ ALAND10            <dbl> 26517362, 26517362, 26517362, 26517362, 26517362, 26517362, 26517362, 26517...
$ AWATER10           <dbl> 52974, 52974, 52974, 52974, 52974, 52974, 52974, 52974, 52974, 52...
$ INTPTLAT10         <chr> "+38.0376579", "+38.0376579", "+38.0376579", "+38.0376579", "+38....
$ INTPTLON10         <chr> "-078.4853806", "-078.4853806", "-078.4853806", "-078.4853806", "...
$ Shape_Leng         <dbl> 47968.96, 47968.96, 47968.96, 47968.96, 47968.96, 47968.96, 47968.96, 47968...
$ Shape_Area         <dbl> 42902561, 42902561, 42902561, 42902561, 42902561, 42902561, 42902561, 42902...
$ geometry           <MULTIPOLYGON [°]> MULTIPOLYGON (((-78.47071 3..., MULTIPOLYGON (((-78...
$ StateAbbreviation <chr> "VA", "VA", "VA", "VA", "VA", "VA", "VA", "VA", "VA", "VA", "VA", "VA",...
$ GeographyName      <chr> "51540(cbwsonly)", "51540(cbwsonly)", "51540(cbwsonly)", "51540(c...
$ Geography          <chr> "Charlottesville City, VA (CBWS Portion Only)", "Charlottesville ...
$ Agency             <chr> "Department of Defense", "Department of Defense", "Department of ...
$ BMPShortName       <chr> "wetpondwetland", "wetpondwetland", "wetpondwetland", "wetpondwet...
$ BMP               <chr> "Wet Ponds and Wetlands", "Wet Ponds and Wetlands", "Wet Ponds an...
$ BMPTYPE            <chr> "Efficiency", "Efficiency", "Efficiency", "Efficiency", "Efficien...
$ Unit               <chr> "Acres Treated", "Acres Treated", "Acres Treated", "Acres Treated...
$ Sector             <chr> "Developed", "Developed", "Developed", "Developed", "Developed", ...
$ FromLoadSource     <chr> "Non-Regulated Roads", "Non-Regulated Buildings and Other", "Non-...
$ ToLoadSource       <chr> "Non-Regulated Roads", "Non-Regulated Buildings and Other", "Non-...
$ AmountSubmitted    <dbl> 0.000000e+00, 0.000000e+00, 0.000000e+00, 0.000000e+00, 0.000000e...
$ AmountBackedOut    <dbl> 0.000000000, 0.000000000, 0.000000000, 0.000000000, 0.000000000, ...
$ AmountNotBackedout <dbl> 0.000000e+00, 0.000000e+00, 0.000000e+00, 0.000000e+00, 0.000000e...
$ AmountCredited     <dbl> 0.000000e+00, 0.000000e+00, 0.000000e+00, 0.000000e+00, 0.000000e...

```

### Task- 2.4

```
> #To solve this question, I got help from online sources.
> distance <- sf::st_distance(dams, streams)
st_as_s2(): dropping Z and/or M coordinate
>
> nearest_distances <- apply(distance, 1, min)
> nearest_stream_index <- apply(distance, 1, which.min)
>
> nearest_features <- data.frame(
+   damID = dams$UNIQUE_ID,
+   streamHYDROID = streams$HYDROID[nearest_stream_index],
+   nearest_stream_index= nearest_stream_index,
+   distance = nearest_distances
+ )
```

```
> print(nearest_features)
```

	damID	streamHYDROID	nearest_stream_index	distance
1	MD_NA001	13098	53	7.821344e-05
2	PA_07-030	246139	2483	0.000000e+00
3	PA_1194387	216524	1711	1.062797e-04
4	PA_12-013	238880	2282	5.349659e-05
5	PA_17-122	220366	1758	0.000000e+00
6	PA_18-037	313466	2667	0.000000e+00
7	PA_31-066	323941	2689	0.000000e+00
8	PA_35-005	318345	2674	0.000000e+00
9	PA_36-019	215867	1700	7.645973e-05
10	PA_36-194	214064	1670	4.775098e-05
11	PA_53-029	219991	1747	7.456252e-05
12	PA_53-030	236000	2192	7.455748e-05
13	PA_54-056	301946	2656	0.000000e+00
14	PA_59-003	249928	2589	0.000000e+00
15	PA_66-022	297817	2653	0.000000e+00
16	PA_PA01006	250092	2590	7.088941e-05
17	VA_15	25539	450	0.000000e+00
18	VA_810	335116	2690	2.975416e+02
19	em20151	87672	1049	0.000000e+00
20	em20152	87674	1050	3.848829e-05
21	em20154	205677	1595	4.565380e-05
22	em20155	277730	2622	0.000000e+00
23	em20156	211091	1626	0.000000e+00
24	em20157	235146	2170	9.029374e-05
25	em20158	125695	1502	0.000000e+00
26	em201513	201064	1592	0.000000e+00
27	em201514	245899	2478	1.074596e-04
28	em201515	252439	2595	7.003663e-05
29	n/a (40-253)	317620	3188	9.873440e-01
30	PA_21-175	316985	3189	6.139861e-01
31	PA_11-074	226155	2908	1.403167e+01
32	PA_40-083	281371	3158	2.921186e+00
33	PA_67-020	213026	3184	7.597703e+00
34	PA_67-021	212887	3186	7.433692e-01

---

## **Task- 2.5**

```
> dams %>% group_by(STATE) %>% summarise(DamsCount=n())
```

Simple feature collection with 3 features and 2 fields  
Geometry type: MULTIPOINT  
Dimension: XY  
Bounding box: xmin: -79.03091 ymin: 37.23352 xmax: -75.53803 ymax: 41.71974  
Geodetic CRS: WGS 84  
# A tibble: 3 × 3

	STATE	DamsCount	geometry
	<chr>	<int>	<MULTIPOINT [°]>
1	MD	2	((-75.73443 38.4618), (-76.06189 39.048))
2	PA	27	((-75.53803 41.46384), (-75.90023 41.20816), (-76.05354 41.6335), (-76.2263 ...
3	VA	5	((-77.41335 37.23352), (-78.89962 38.05994), (-78.89175 38.05944), (-78.0075...

---