A-PLT Day 2

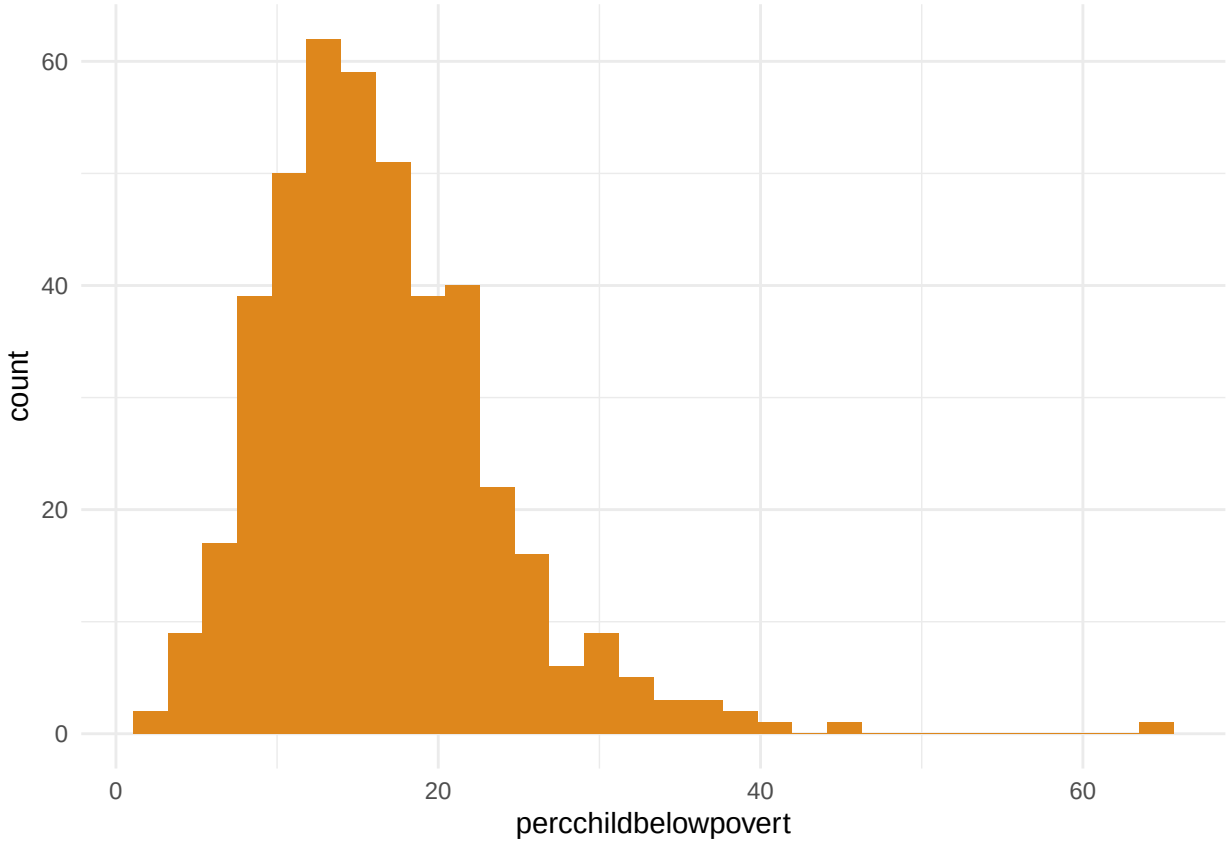
Haseeb Raza

2024-07-10

Data distribution through graphs

Histogram can be used to check the normality of data

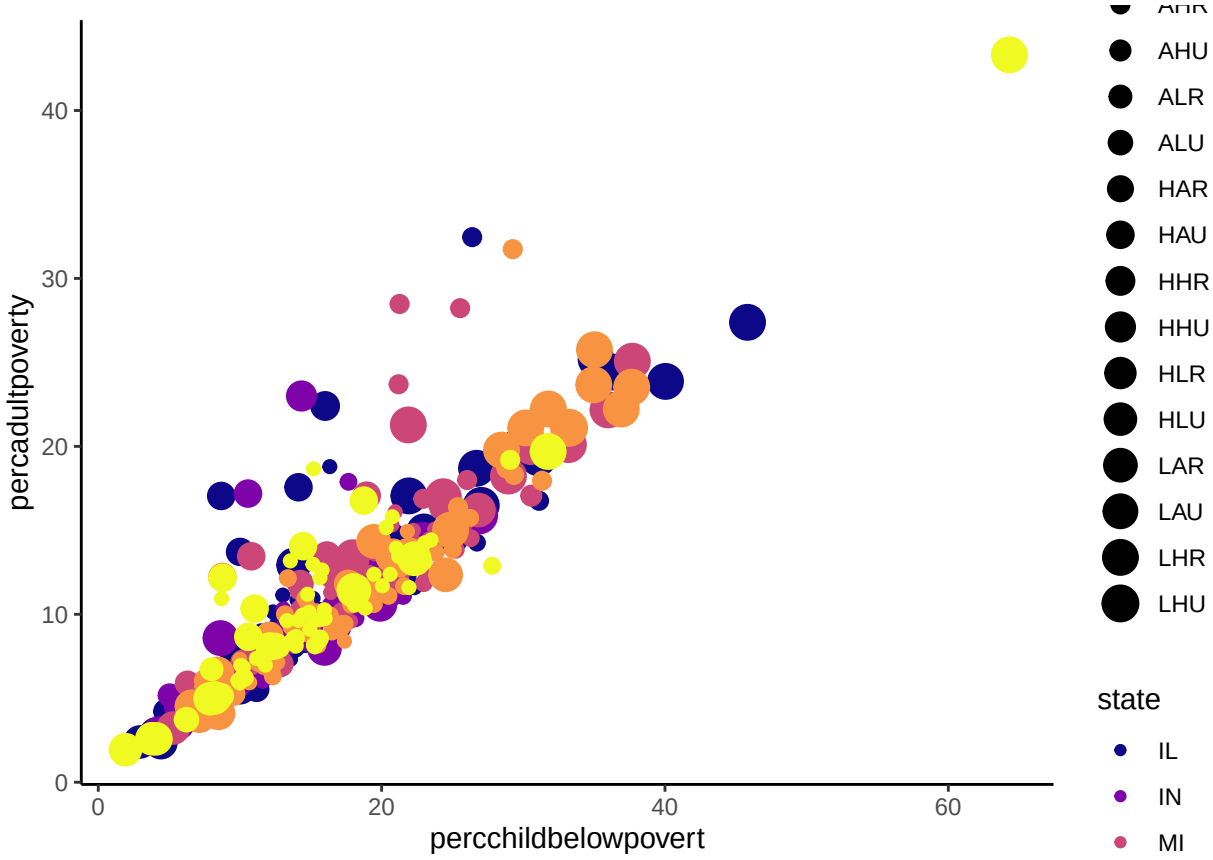
```
library(ggplot2)
ggplot(midwest) +
  aes(x = percchildbelowpovert) +
  geom_histogram(bins = 30L, fill = "#DE871C") +
  theme_minimal()
```



Scatterplot is used to visualize the distribution of two variables

```
library(ggplot2)
ggplot(midwest) +
  aes(
    x = percchildbelowpovert,
    y = percadultpoverty,
    colour = state,
    size = category
  ) +
  geom_point() +
  scale_color_viridis_d(option = "plasma", direction = 1) +
  theme_classic()
```

```
## Warning: Using size for a discrete variable is not advised.
```



Graphs can be merged for a better understanding

```
library(ggplot2)
ggplot(midwest) +
 aes(x = percchildbelowpovert, y = percadultpoverty) +
```

```
geom_point(colour = "#112446") +
geom_smooth(aes(fill = state), se = TRUE, colour = "#112446") +
scale_fill_viridis_d(option = "viridis",
direction = 1) +
labs(x = "PCP", y = "PAP", title = "Linear Regression", subtitle = "under YPDC's Peer Mentorship") +
theme_classic() +
theme(legend.text = element_text(face = "bold", size = 12L), legend.title = element_text(face = "bold"
size = 14L))
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```