Data Visualization with ggplot2

Evan L. Ray

February 02, 2018
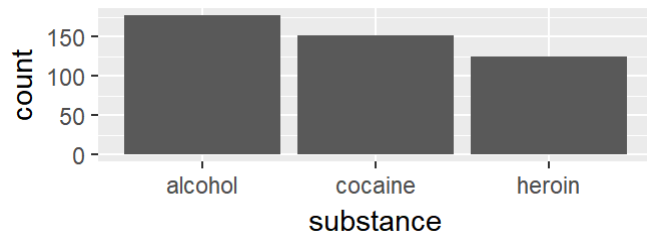
Warm Up (with a neighbor)

Here are the first few rows of a data set with information about participants in a randomized controlled trial designed to evaluate a substance abuse treatment program. What are the observational units and variables? Are the variables categorical or quantitative?

```
head(HELPrct)
```
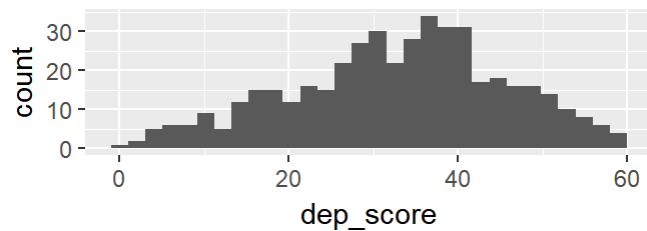
```
##    age homeless substance dep_score mental_score
## 1  37   housed   cocaine        49    25.111990
## 2  37 homeless   alcohol        30    26.670307
## 3  26   housed    heroin        39     6.762923
## 4  39   housed    heroin        15    43.967880
## 5  32 homeless   cocaine        39    21.675755
## 6  47   housed   cocaine         6    55.508991
```
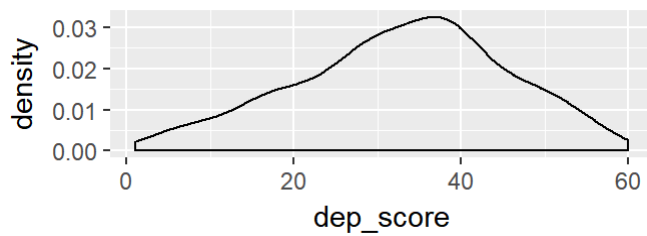
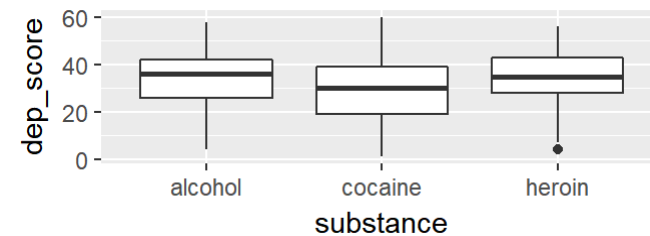# 5 Main Plots (for this class)

1. Bar: Categorical
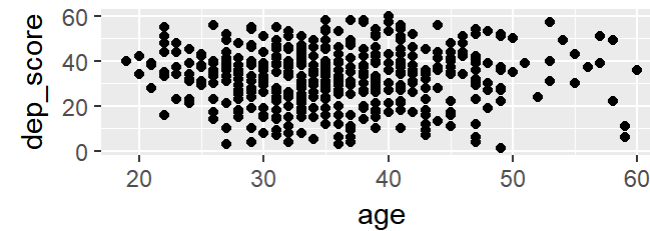

4. Boxplot: 1 Quantitative, 1 Categorical

2. Histogram: Quantitative

5. Scatterplot: 2 Quantitative
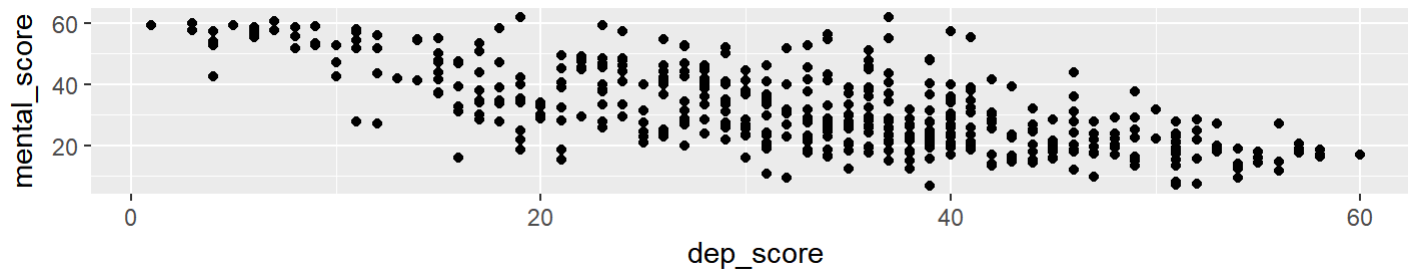

3. Density: Quantitative

Also: color

The Grammar of Graphics

A statistical graphic is a `mapping` of `data` variables to `aes()`thetic attributes of `geom_`etric objects

```
ggplot(data = <data_frame>,
    mapping = aes(<attribute1> = <variable1>, <attribute2> = <variable2>),
 ) +
 geom_<geometry type>()
```
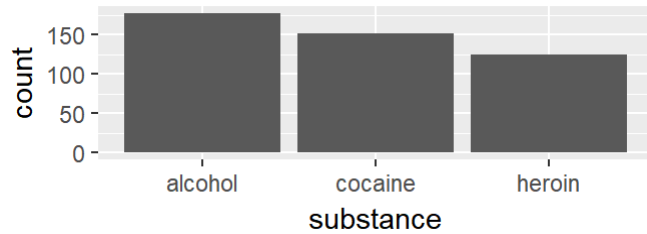
```
ggplot(data = HELPrct,
    mapping = aes(x = dep_score, y = mental_score),
 ) +
 geom_point()
```
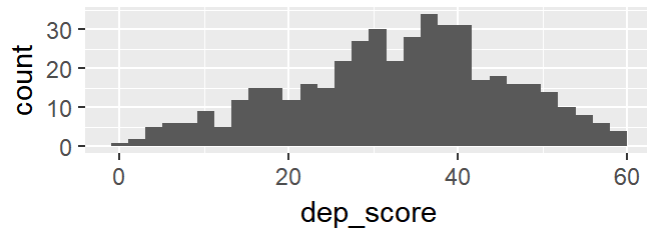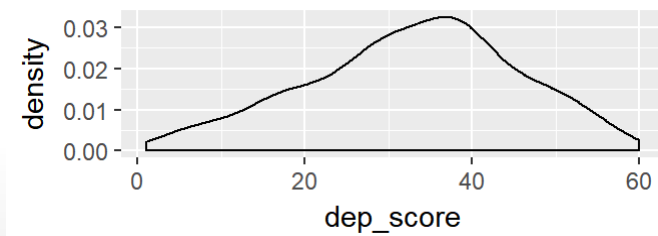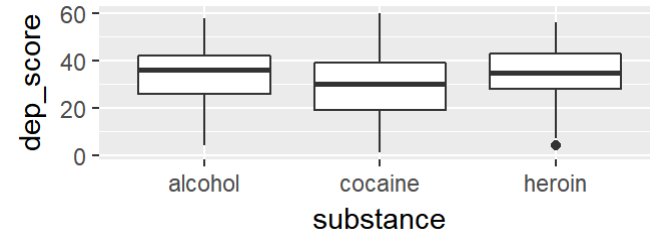
What are the **geom_**etric objects?

1. geom_bar

4. geom_boxplot

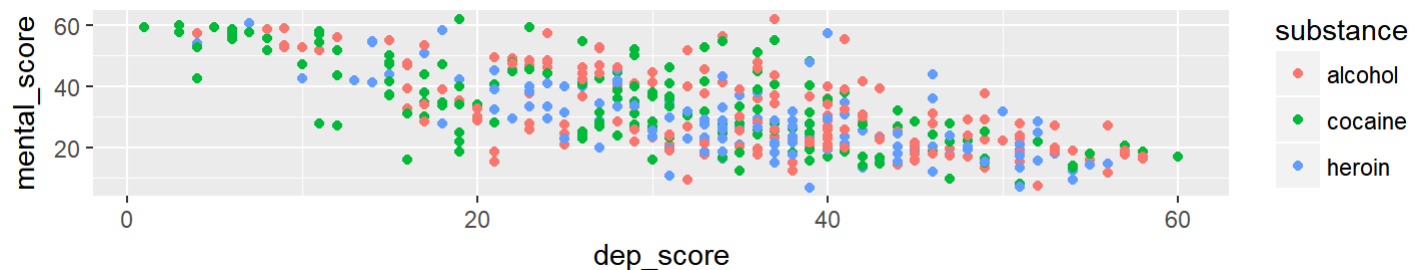2. geom_histogram

5. geom_point
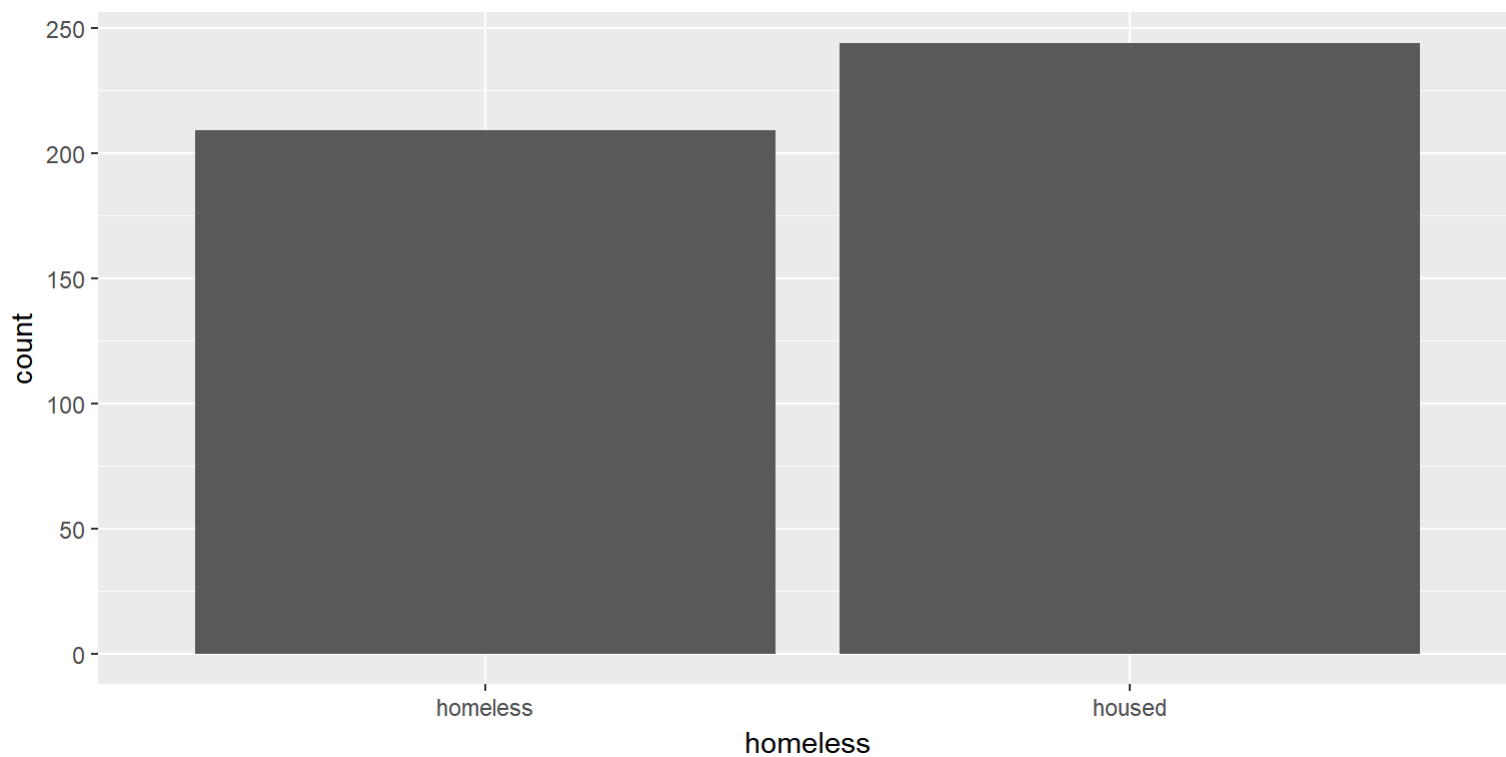
3. geom_density

What are the **aes**thetic properties?

1. `x` for horizontal axis (applicable for all geometry types)

2. `y` for vertical axis (applicable for plot types that have a variable on the vertical axis: boxplot and scatter plot)

3. `fill` (for fill color in geometric objects that have an internal area) or `color` (for color of objects that don't have an internal area)

```
ggplot(data = HELPrct,
    mapping = aes(x = dep_score, y = mental_score, color = substance),
  ) +
  geom_point()
```
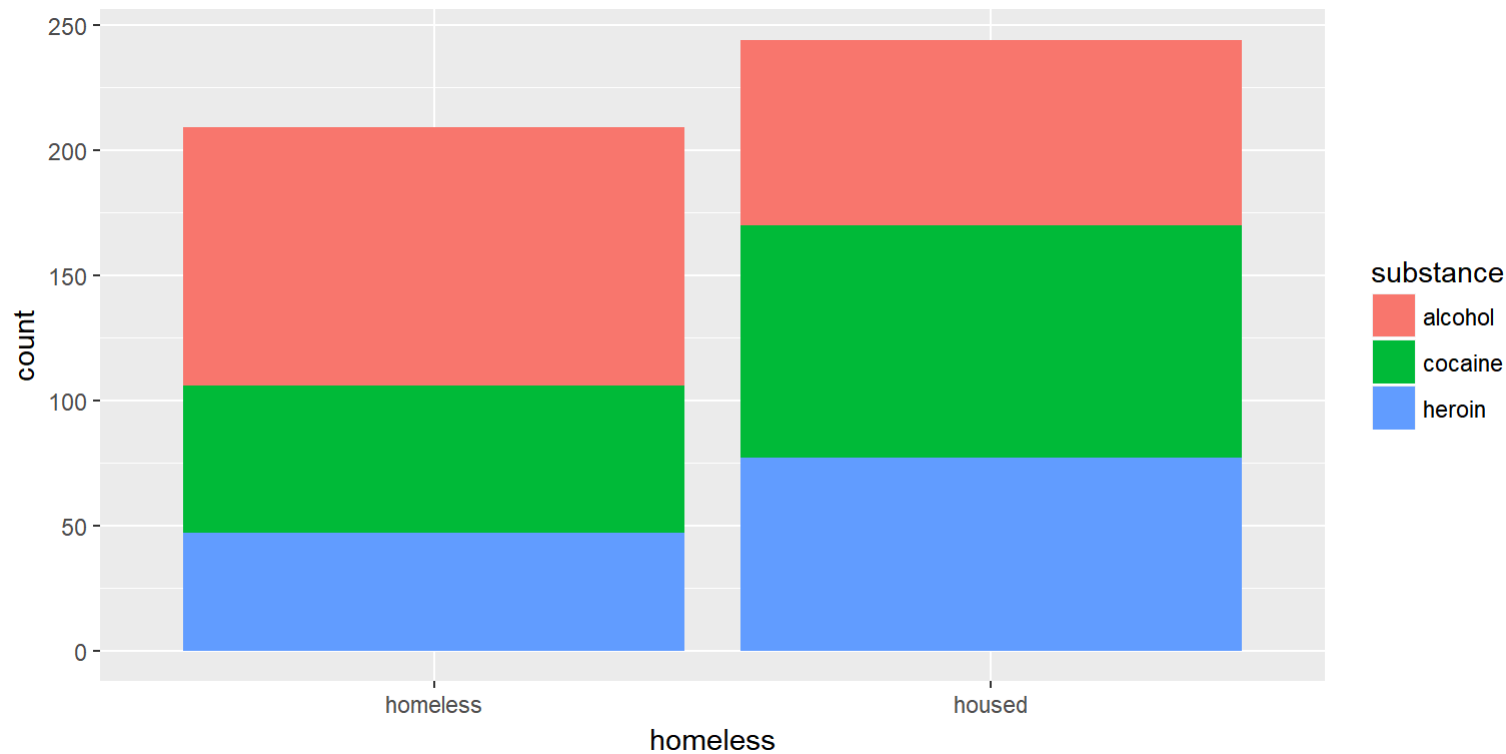
1 Categorical Variable: Bar Plot

```
ggplot(data = HELPrct,
    mapping = aes(x = homeless)
  ) +
  geom_bar()
```
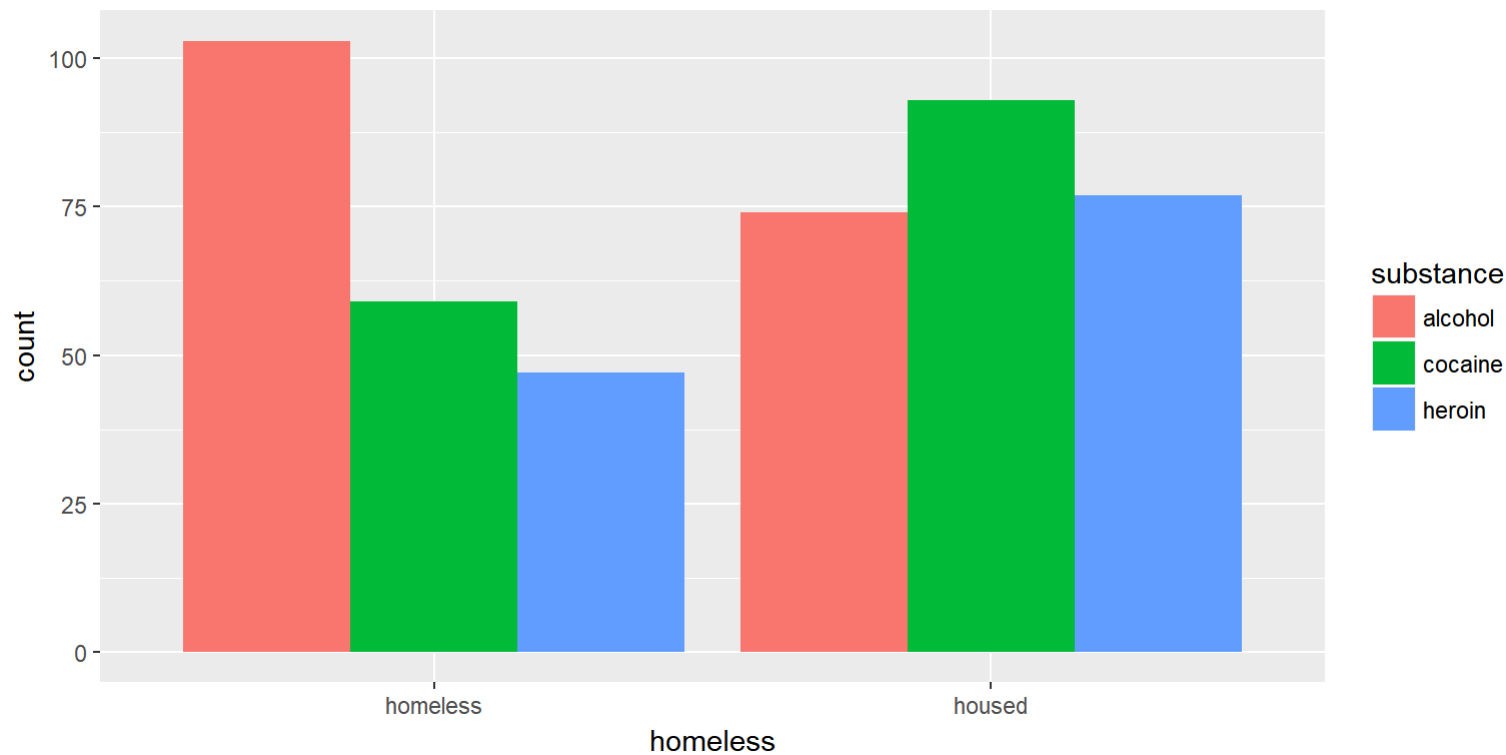
2 Categorical Variables: Bar Plot

```
ggplot(data = HELPrct,
    mapping = aes(x = homeless, fill = substance)
  ) +
  geom_bar()
```
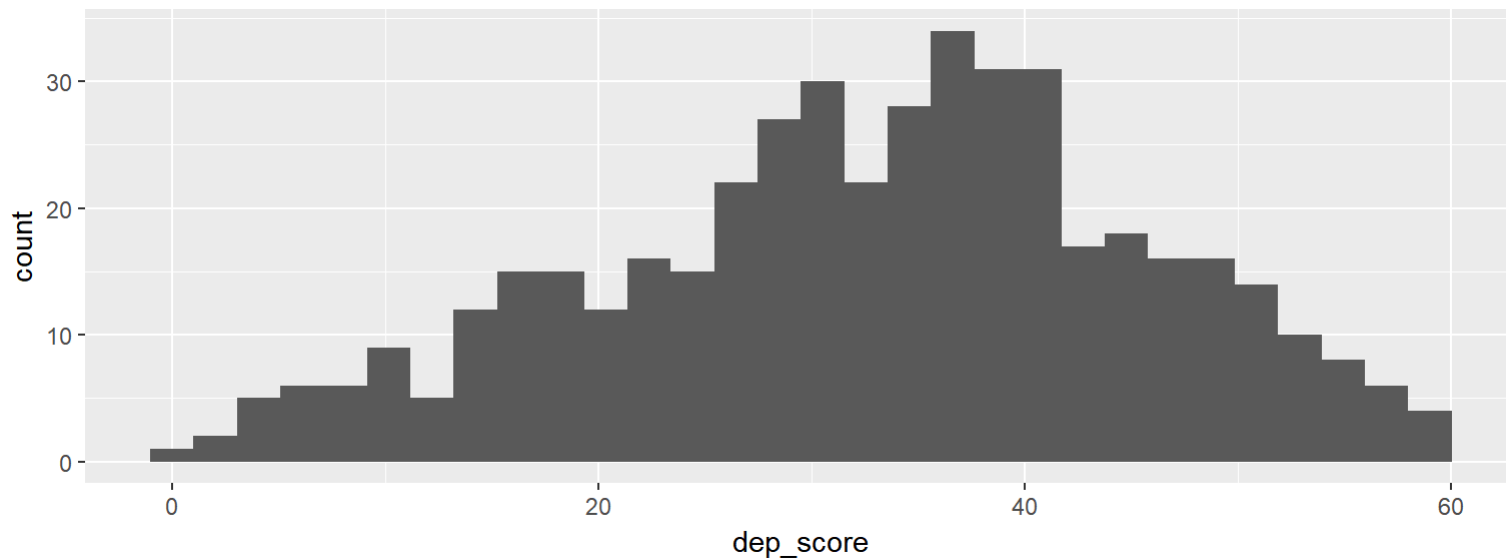
2 Categorical Variables: Bar Plot

```
ggplot(data = HELPrct,
    mapping = aes(x = homeless, fill = substance)
) +
  geom_bar(position = position_dodge())
```

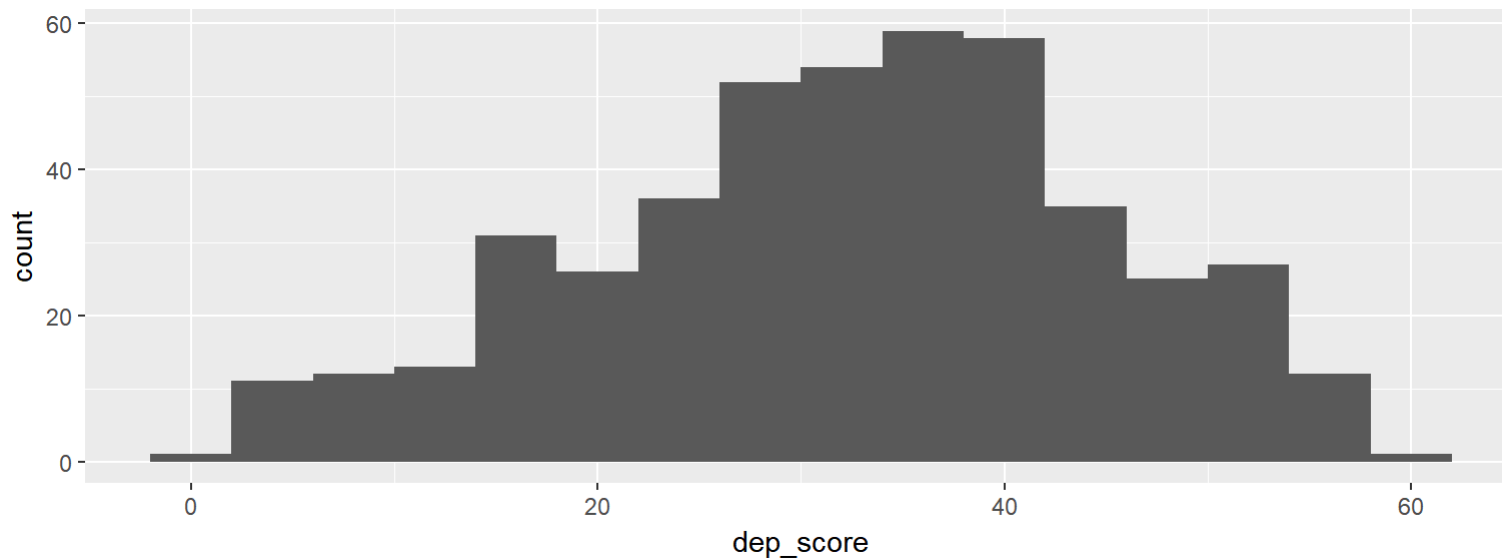# 1 Quantitative Variable: Histograms

```
ggplot(data = HELPrct,
    mapping = aes(x = dep_score),
  ) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
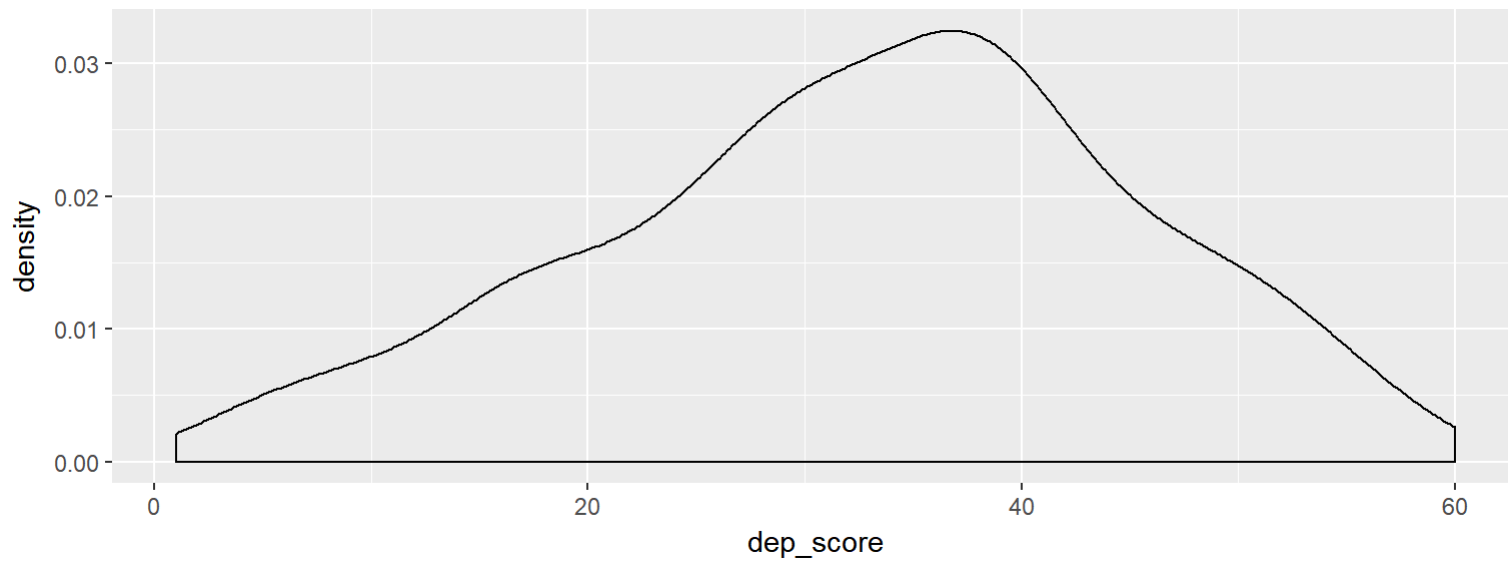
1 Quantitative Variable: Histograms

```r
ggplot(data = HELPrct,
    mapping = aes(x = dep_score),
  ) +
  geom_histogram(binwidth = 4)
```

1 Quantitative Variable: Density Plots

```
ggplot(data = HELPrct,
    mapping = aes(x = dep_score),
  ) +
  geom_density()
```

2 Quantitative Variables: Scatter Plots

```
ggplot(data = HELPrct,
    mapping = aes(x = dep_score, y = mental_score)
  ) +
  geom_point()
```
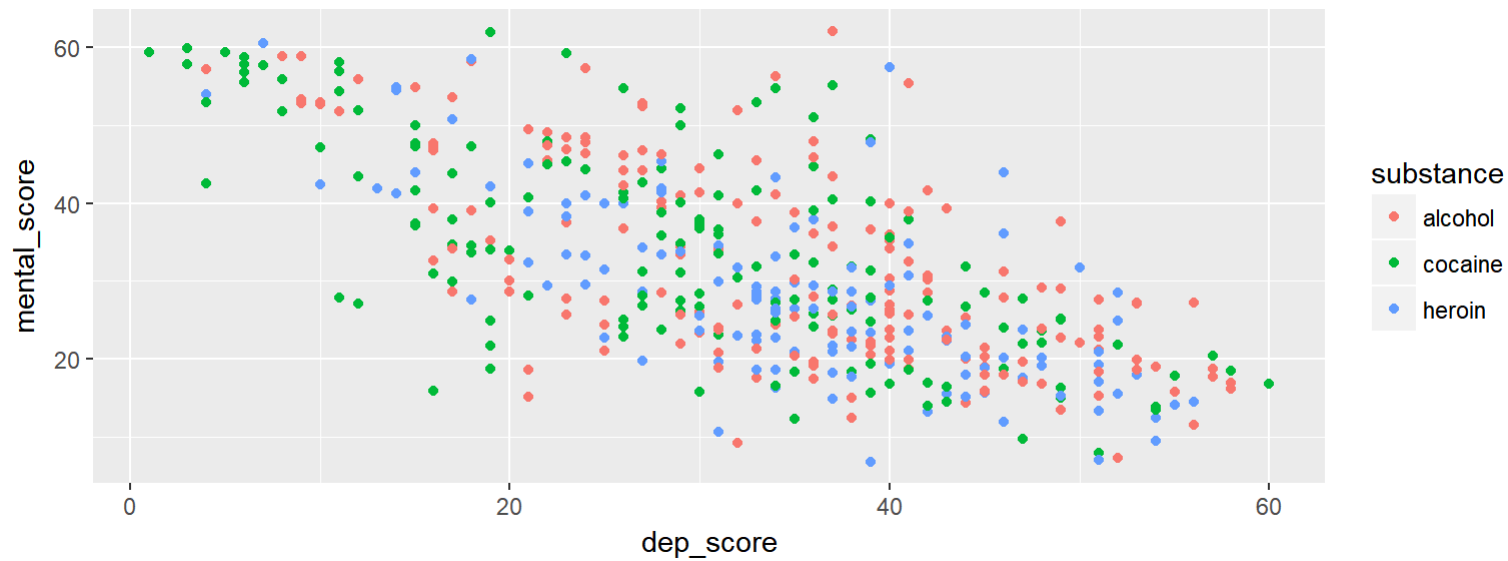
2 Quantitative Variables: Scatter Plots

```
ggplot(data = HELPrct,
    mapping = aes(x = dep_score, y = mental_score, color = substance)
) +
geom_point()
```

1 Quantitative, 1 Categorical: Box Plots

```
ggplot(data = HELPrct,
    mapping = aes(x = substance, y = dep_score)
  ) +
  geom_boxplot()
```
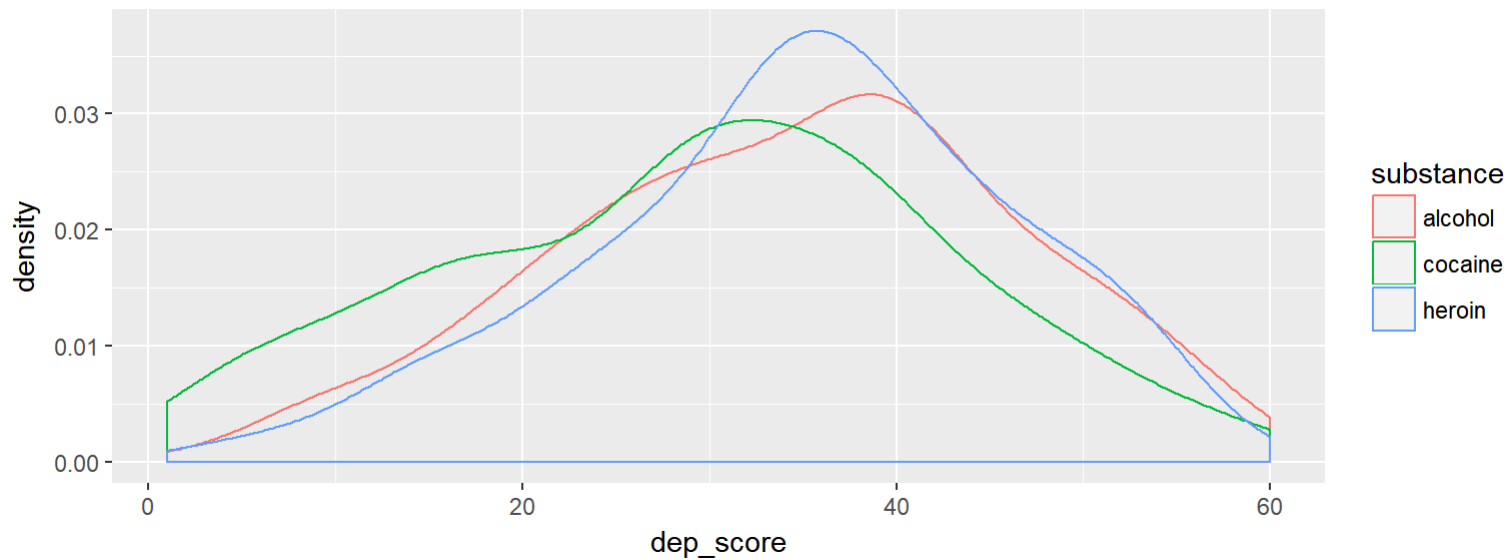
1 Quantitative, 1 Categorical: Density

```
ggplot(data = HELPrct,
    mapping = aes(x = dep_score, color = substance)
  ) +
  geom_density()
```
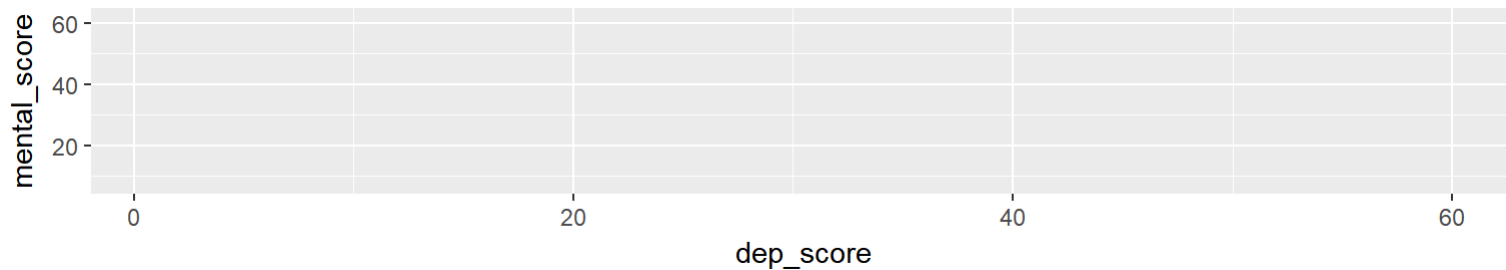
Summary

Variables	Plot Type	Geometry	Aesthetics
1 Categorical	Bar Plot	geom_bar	x
1 Quantitative	Histogram	geom_histogram	x
1 Quantitative	Density	geom_density	x
2 Categorical	Bar Plot	geom_bar	x, fill
2 Quantitative	Scatter Plot	geom_point	x, y
1 Categorical and 1 Quantitative	Box Plot	geom_boxplot	x (categorical), y (quantitative)
1 Categorical and 1 Quantitative	Density Plot	geom_density	x (quantitative), color (categorical)

Common Error 1

No **+** after call to `ggplot()`:

```
ggplot(data = HELPrct,
    mapping = aes(x = dep_score, y = mental_score)
  )
```



```
  geom_point()
```

```
## geom_point: na.rm = FALSE
## stat_identity: na.rm = FALSE
## position_identity
```

Common Error 2

%>% instead of **+** after call to `ggplot()`:

```
ggplot(data = HELPrct,
    mapping = aes(x = dep_score, y = mental_score)
  ) %>%
  geom_point()
```

Error: Mapping must be created by **aes()** or **aes_()**

Common Error 3

Forgot `data =`:

```
ggplot(mapping = aes(x = dep_score, y = mental_score)) +
  geom_point()
```

Error in FUN(X[[i]], ...) : object 'dep_score' not found