



DATA VISUALIZATION  
WITH



PRINCIPLES & PRACTICE



**secret behind the smile!**

# HELLO!

I am **Elijah Appiah** from **Ghana**.

I am an Economist by profession.

I love everything data, so I love R!

You can reach me:

**[eappiah.uew@gmail.com](mailto:eappiah.uew@gmail.com)**

## Lesson Goals

- ▶ Provide compact introduction to allow readers learn about visualization techniques.
- ▶ Emphasize the strong connections between visualizations and insight.

# Datasets

- ▶ **mtcars**: base R
- ▶ **wage1**: wooldridge package
- ▶ **diamonds**: ggplot2 package

# Variables

## Categorical

**Nominal** – *names, labels, categories with no natural order*

**e.g.** gender, countries

**Ordinal** – *categories with an order*

**e.g.** Likert Scales

## Numeric

**Discrete** – *counts*

**e.g.** number of cylinders of a vehicle

**Continuous** – *measured even within an interval*

**e.g.** height, weight

# Variables (e.m)

- ▶ **Discrete** – represents counts  
e.g. number of students, grade levels, gender,  
number of blue marbles in a jar, etc.
- ▶ **Continuous** – represents measurable  
amounts  
e.g. height, weight, temperature, distance, etc.

7

# GGPLOT2

**GRAMMAR OF GRAPHICS PLOTS**

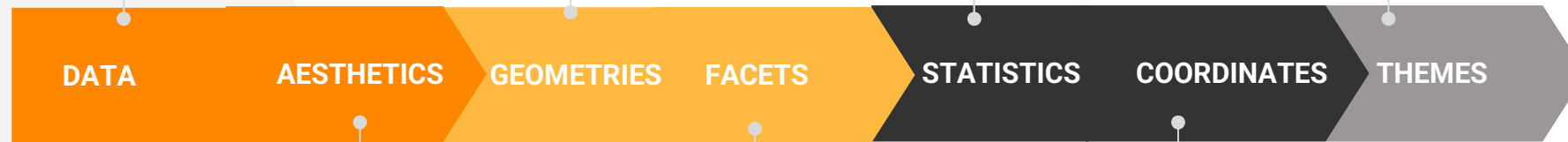
# GGPLOT2 LAYERS

Dataset to be  
visualized

Visual elements  
for the data

Data representations  
to aid understanding

Plot appearance  
(all non-data ink)



Scales onto  
which data is  
mapped

Create multiple  
plots

The space on  
which data is  
plotted



## GGPLOT2 LAYERS

Dataset to be  
visualized

**DATA**

Visual elements  
for the data

**AESTHETICS**

**GEOMETRIES**

Scales onto  
which data is  
mapped



# GGPLOT2

- ▶ The package is


`ggplot2`

- ▶ The function is

`ggplot()`

# Layer: DATA

```
ggplot(data = df)
```



*Blank canvas with  
grey background*

# Layer: AESTHETICS

**The aesthetic attributes include:**

**x, y, colour (or color), shape, size, fill, alpha,  
etc...**

- ▶ Aesthetics are mapped in the **aes()** function in the **ggplot()** function.

## Layer: AESTHETICS

```
ggplot(data = df, mapping = aes())
```

**Aesthetic  
attributes**



# Layer: AESTHETICS

```
ggplot(data = df, aes())
```

**Aesthetic  
attributes**



# Layer: AESTHETICS

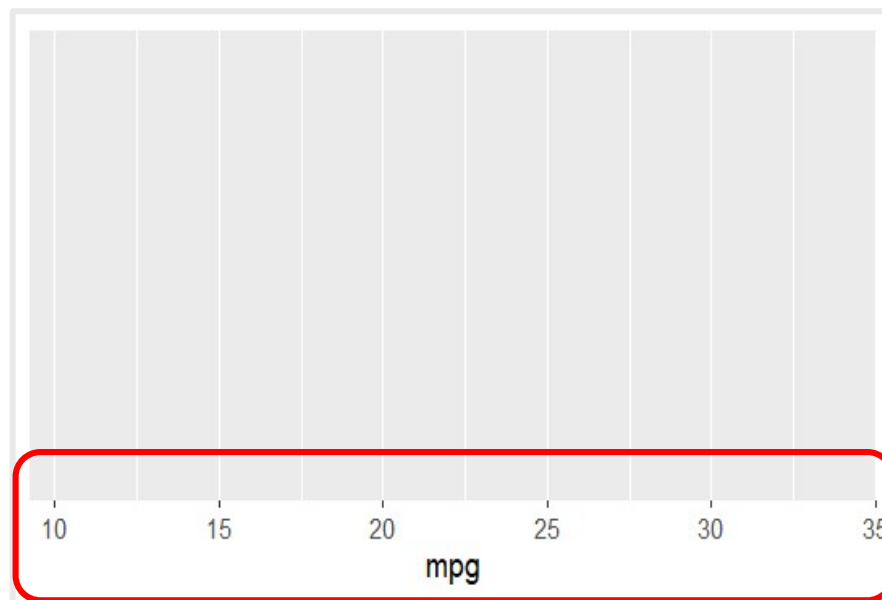
```
ggplot(df, aes())
```

**Aesthetic  
attributes**



# Layer: AESTHETICS

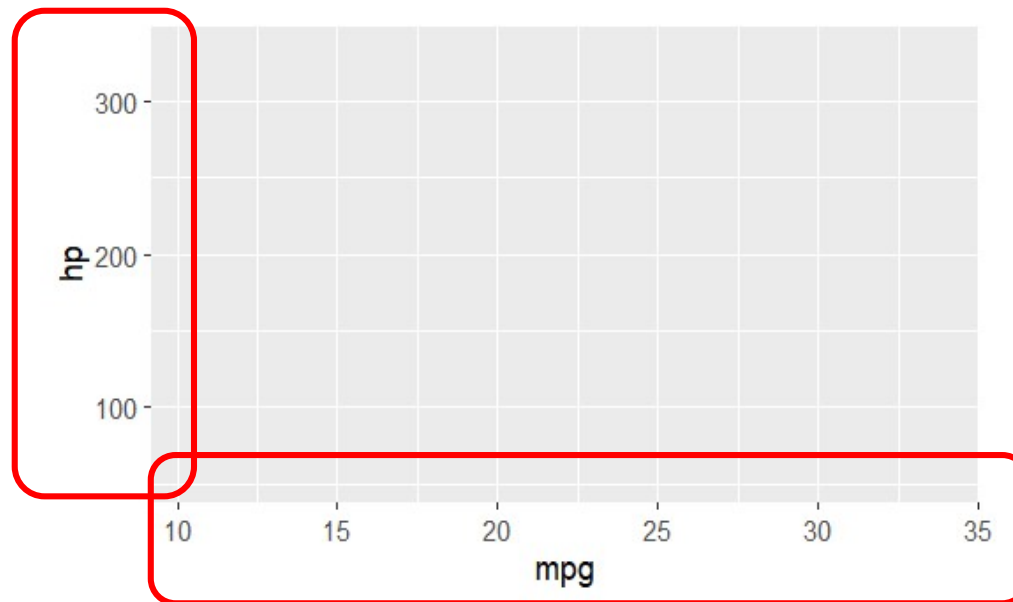
```
ggplot(mtcars, aes(x = mpg))
```





# Layer: AESTHETICS

```
ggplot(mtcars, aes(x = mpg, y = hp))
```



# Layer: GEOMETRIES

- ▶ The visual elements of plots are defined by **geoms**.
- ▶ It is specified as **geom\_\*()**.  
where \* denotes the specific **type of plot** to create.
- ▶ A bar plot will be **geom\_bar()**
- ▶ A histogram will be **geom\_histogram()**
- ▶ A scatter plot will be **geom\_point()**

*Don't worry.....we will be going into details soon.....*

# Layer: GEOMETRIES

- ▶ The geometric objects (or geoms) are **added** (+) to the `ggplot()` function.

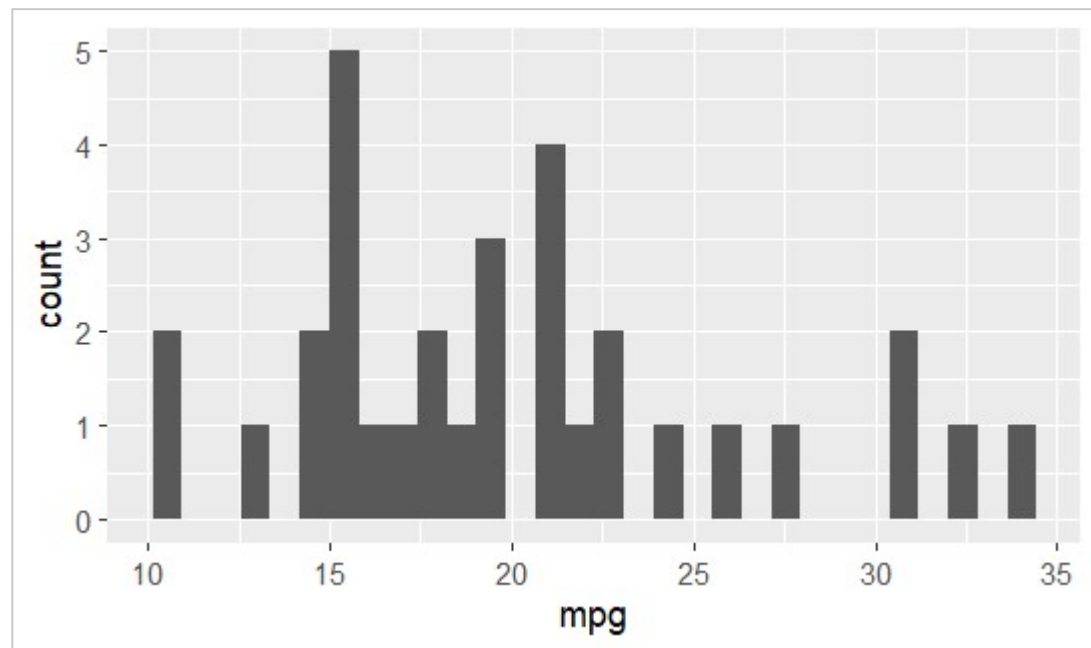
## Example:

```
ggplot(mtcars, aes(x = mpg)) +  
  geom_histogram()
```

```
ggplot(mtcars, aes(x = mpg, y = hp)) +  
  geom_point()
```

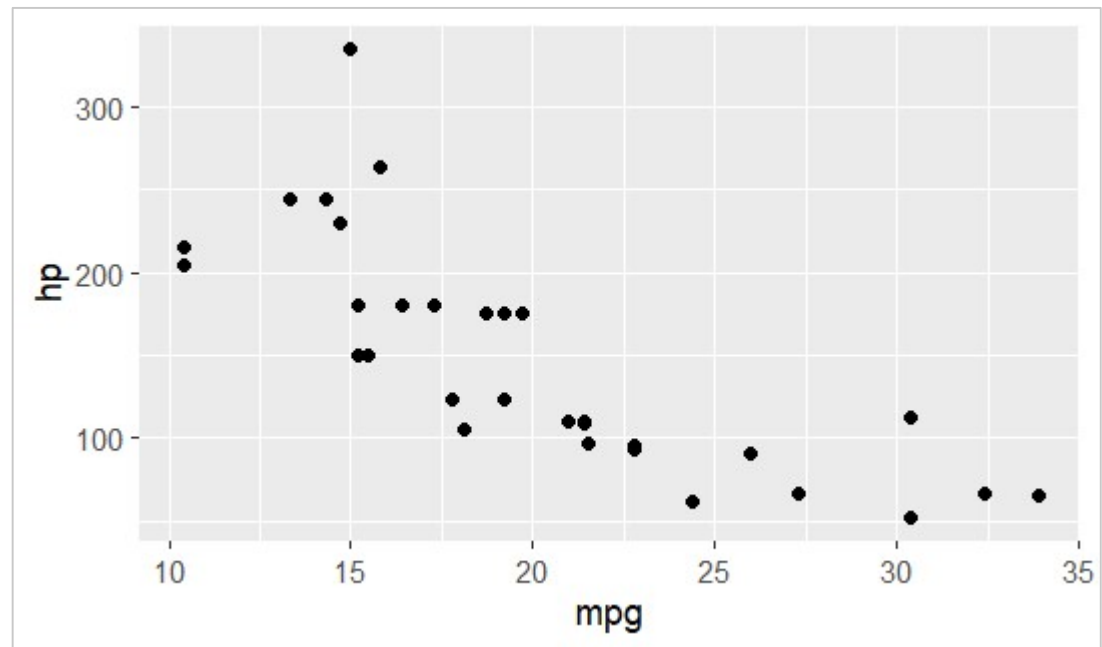
# Layer: GEOMETRIES

```
ggplot(mtcars, aes(x = mpg)) + geom_histogram()
```



# Layer: GEOMETRIES

```
ggplot(mtcars, aes(x = mpg, y = hp)) + geom_point()
```





Now, let's practice

# VARIABLES & PLOTS - GEOMS

## ▶ ONE VARIABLE

Discrete	Continuous
Bar Plot – <code>geom_bar()</code>	Histogram – <code>geom_histogram()</code>
	Density Plot – <code>geom_density()</code>
	Dot plot – <code>geom_dot()</code>
	Frequency Polygons – <code>geom_freqpoly()</code>

# VARIABLES & PLOTS – GEOMS

- ▶ `geom_bar()` – display distribution of discrete variables.
- ▶ `geom_histogram()` – bin and count continuous variable, display with bars.
- ▶ `geom_density()` – smoothed density estimate.
- ▶ `geom_dotplot()` – stack individual points into a dot plot.
- ▶ `geom_freqpoly()` – bin and count continuous variable, display with lines.





Now, let's practice

# VARIABLES & PLOTS - GEOMS

## ▶ TWO VARIABLES

Both Continuous	One Continuous, One Discrete
Scatter plot – <code>geom_point()</code>	Bar plot – <code>geom_col()</code> or <code>geom_bar(stat="identity")</code>
Quantile plot – <code>geom_quantile()</code>	Box Plot – <code>geom_boxplot()</code>
Rug plot – <code>geom_rug()</code>	Violin plot – <code>geom_violin()</code>
Text labels – <code>geom_text()</code>	

# VARIABLES & PLOTS – GEOMS

- ▶ `geom_point()` – scatterplot.
- ▶ `geom_quantile()` – smoothed quantile regression.
- ▶ `geom_rug()` – marginal rug plots.
- ▶ `geom_text()` – text labels.
- ▶ `geom_col()/geom_bar(stat="identity")` – bar chart of precomputed summaries.
- ▶ `geom_boxplot()` – boxplots.
- ▶ `geom_violin()` – show density of values in each group.



Now, let's practice

# VARIABLES & PLOTS - GEOMS

## ▶ TWO VARIABLES

At Least One Discrete	Show Distribution (continuous)
Count plot– <code>geom_count()</code>	Hexagonal Heatmap – <code>geom_hex()</code>
Jitter plot– <code>geom_jitter()</code>	Heatmap – <code>geom_bin2d()</code>
	Density plot – <code>geom_density2d()</code>

# VARIABLES & PLOTS – GEOMS

- ▶ `geom_count()` – count number of points at distinct locations.
- ▶ `geom_jitter()` – randomly jitter overlapping points.
- ▶ `geom_hex()` – bin into hexagons and count.
- ▶ `geom_bin2d()` – smoothed 2d density estimate.
- ▶ `geom_density2d()` – smoothed 2d density estimate.



Now, let's practice

# VARIABLES & PLOTS - GEOMS

## ▶ TWO VARIABLES

One Time, One Continuous	Display Uncertainty
Line plot – <code>geom_line()</code>	<code>geom_crossbar()</code>
Area plot – <code>geom_area()</code>	<code>geom_errorbar()</code>
Step plot – <code>geom_step()</code>	<code>geom_linerange()</code>
	<code>geom_pointrange()</code>



# VARIABLES & PLOTS - GEOMS

- ▶ `geom_line()` – line plot.
- ▶ `geom_area()` – area plot.
- ▶ `geom_step()` – step plot.
- ▶ `geom_crossbar()` – vertical bar with center.
- ▶ `geom_errorbar()` – error bars.
- ▶ `Geom_linerange()` – vertical line.
- ▶ `geom_pointrange()` – vertical line with center.



Now, let's practice

# VARIABLES & PLOTS - GEOMS

- ▶ **TWO VARIABLES**

`geom_map()` – for map data

- ▶ **THREE VARIABLES**

`geom_contour()` – contours.

`geom_tile()` – tile the plane with rectangles.

`geom_raster()` – equal sized tiles (fast version of `geom_tile()`)

# THANKS!

**Any questions?**

You can find me at: **[eappiah.uew@gmail.com](mailto:eappiah.uew@gmail.com)**