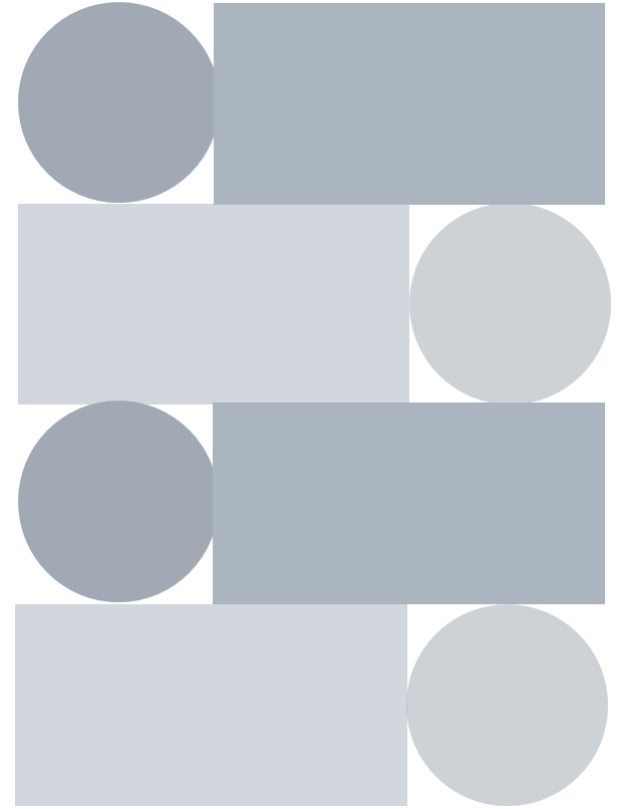
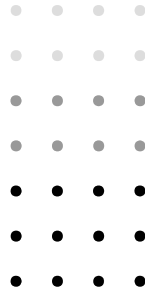


# STATISTICAL COMPUTATION

## WEEK 2 – DESCRIPTIVE STATISTICS

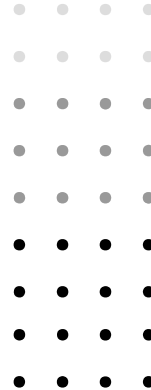
Annisa Auliya  
I Melda Puspita



# GET TO KNOW US

ANNISA AULIYA R.

082334174749

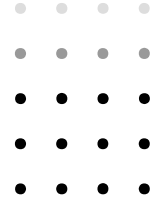


I MELDA PUSPITA L.

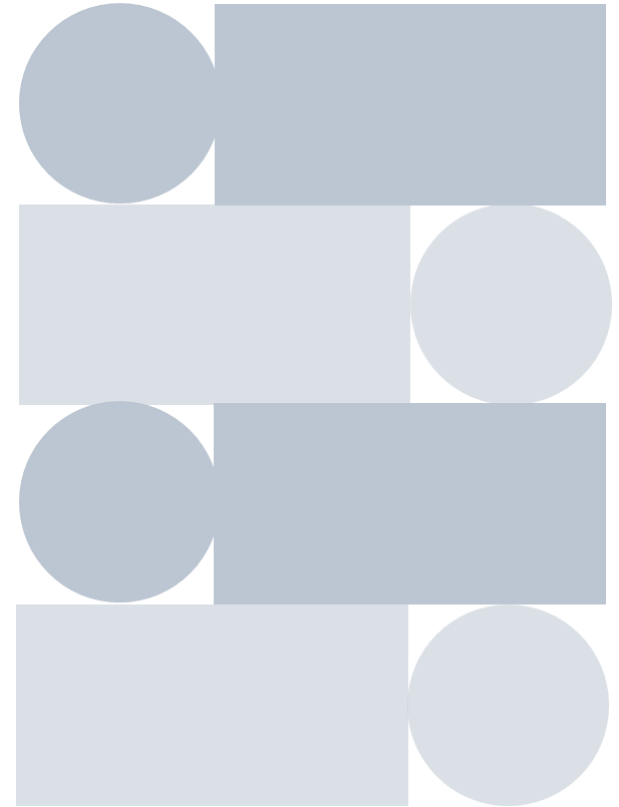
085257113961

<https://intip.in/KomstatC2023>

# MATERIALS



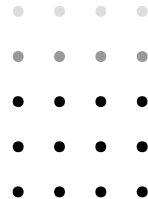
- Conditional Statement & Looping
- Data Manipulation
- Descriptive Statistics



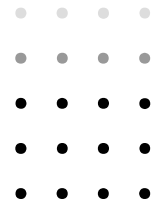
# 01

## CONDITIONAL STATEMENT & LOOPING

---



# STATEMENT



## IF

```
x=1
if (x==2)
{
print ("x=2")
}
```

## IF-ELSE

```
x=1
if (x==2)
{
print ("x = 2")
}
else
{
print ("x != 2")
}
```

## FOR

```
for (i in 1:4)
{
print (i)
}
[1] 1
[1] 2
[1] 3
[1] 4
```

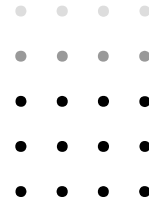
## WHILE

```
i=0
while (i<4)
{
i=i+1
print (i)
}
[1] 1
[1] 2
[1] 3
[1] 4
```

## REPEAT

```
i=0
repeat
{
i=i+1
print (i)
if (i==4) break
}
[1] 1
[1] 2
[1] 3
[1] 4
```

# COMMAND



## APPLY

```
matrix (1:10 , nrow =2) -> a
apply (a ,1 , mean )
[1] 5 6
apply (a ,2 , mean )
[1] 1.5 3.5 5.5 7.5 9.5
```

## SAPPLY & LAPPLY

```
matrix (2:11 , nrow =2) -> a
matrix (1:10 , nrow =2) -> b
c <- list (a,b)
sapply (c, mean )
[1] 6.5 5.5
lapply (c, mean )
[[1]]
[1] 6.5 7 8
[[2]]
[1] 5.5
```

## TAPPLY

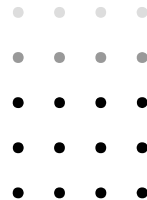
```
data(iris)
attach(iris)
tapply(Sepal.Width,Species,mean)
setosa      versicolor  virginica
3.428      2.770      2.974
```

# 02

# DATA MANIPULATION

---

Insert, Drop, Filter



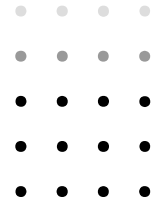


# ACCESS ELEMENT

```
myframe[,1]
myframe["Age "]
myframe$Age
myframe[3,3]=2 # change value
myframe[, -2] # access all column except column 2
attach(myframe) #add obj to search path
Sex
detach("myframe") #del obj from search path
```

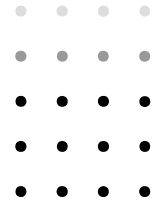


# SUBGROUP



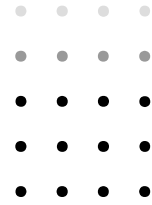
```
subset(myframe,myframe$Age>30) #4 entries
mean(subset(myframe$Age,myframe$Sex=="m"))
mean(subset(myframe$Age,myframe$Sex=="f"))
myframe[(myframe$Sex=="m")&(myframe$Age>30),] #males over
30myframe[(myframe$Sex=="m")|(myframe$Age>30),] #male or over 30
```

# ADD/DELETE COLUMN



```
myframe=cbind(myframe,"Income"=c(1700,2100,2300,2050,2800,1450,3400,2000)  
) #add column  
myframe$Income=NULL #del col
```

# SORTING

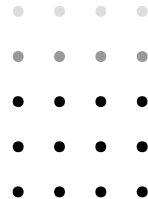


```
> x = c(2,3,5,2,5,6,7,3)
> sort(x)
[1] 2 2 3 3 5 5 6 7
> order(x)
[1] 1 4 2 8 3 5 6 7
> rank(x)
[1] 1.5 3.5 5.5 1.5 5.5 7.0 8.0 3.5
myframe[order(myframe$Sex, partial=myframe$Age),]
```

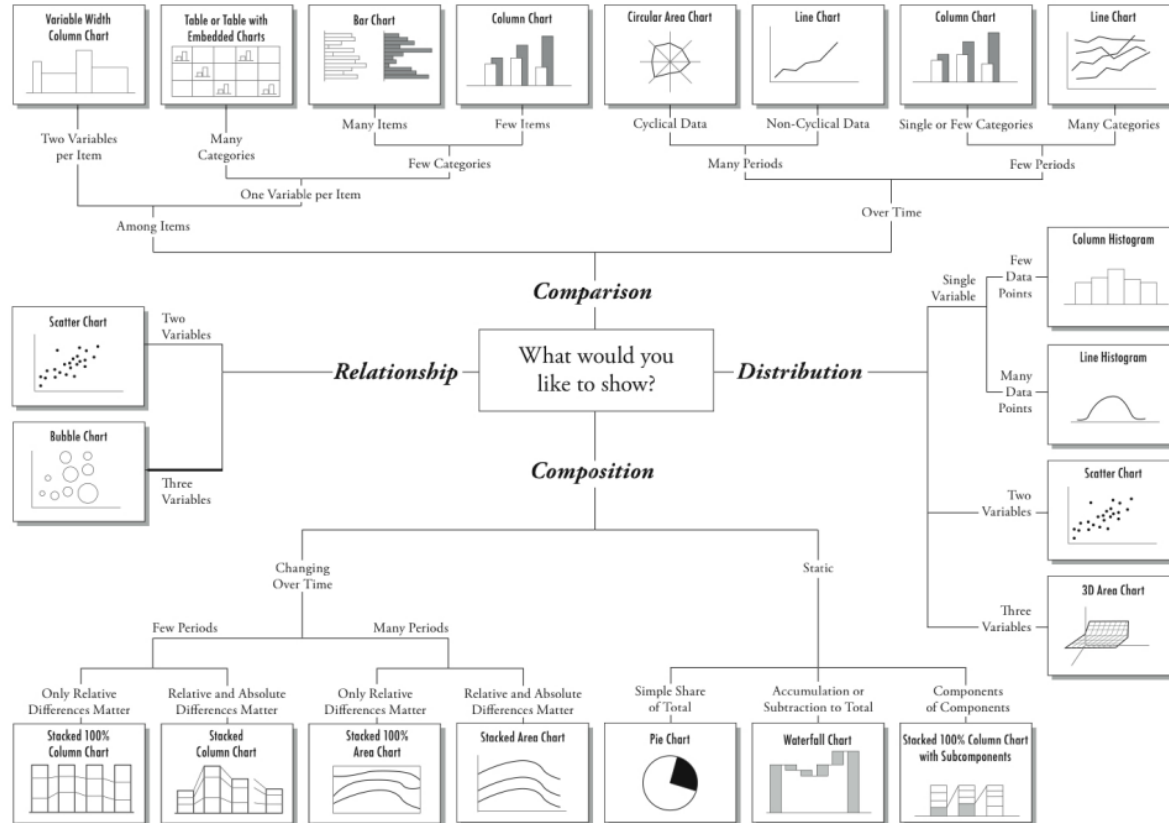
# 03

## DESCRIPTIVE STATISTICS

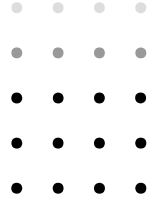
---



# VISUALIZATION



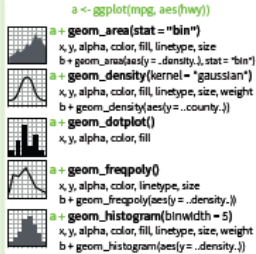
# VISUALIZATION



**Geoms** - Use a geom to represent data points, use the geom's aesthetic properties to represent variables. Each function returns a layer.

## One Variable

### Continuous



### Discrete

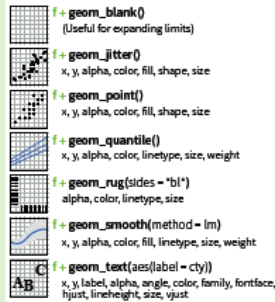


## Graphical Primitives

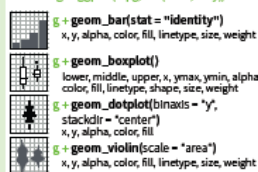


## Two Variables

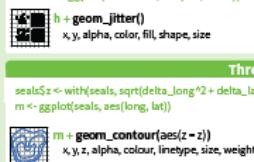
### Continuous X, Continuous Y



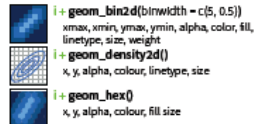
### Discrete X, Continuous Y



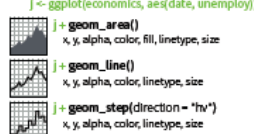
### Discrete X, Discrete Y



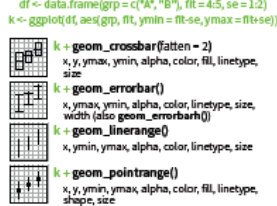
### Continuous Bivariate Distribution



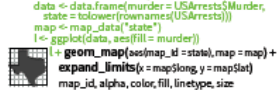
### Continuous Function



### Visualizing error



### Maps



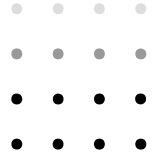
## Three Variables



Find More:

<https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-visualization.pdf>

# VISUALIZATION

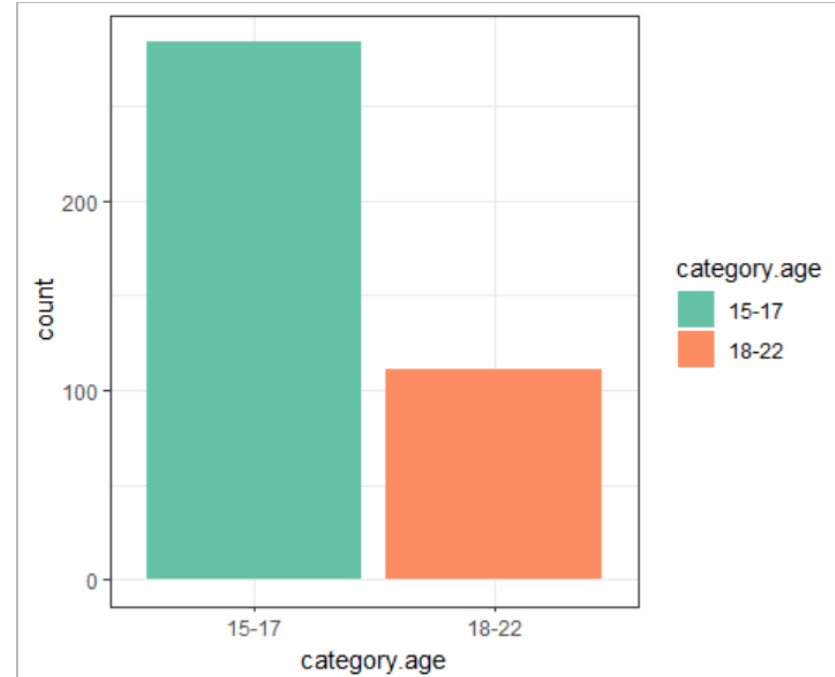


## Bar Chart

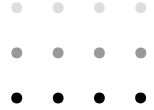
```
library(ggplot2)
ggplot(data,aes(x=category.age,fill=category.age))+
  geom_bar()+theme_bw()+
  scale_fill_brewer(palette="Set2")
```

### Find More:

<http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualization>



# VISUALIZATION

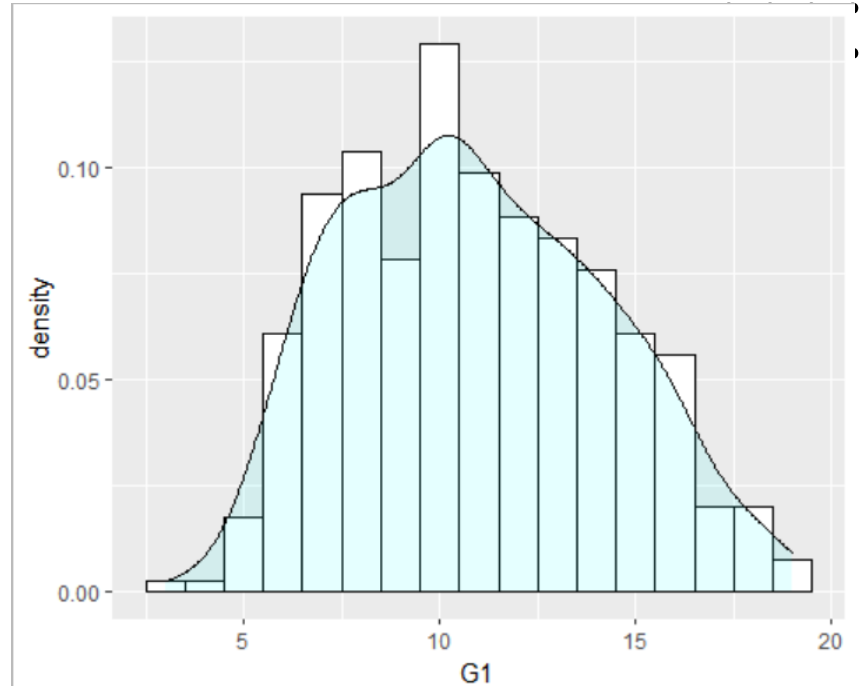


## Histogram

```
ggplot(data,aes(x=G1))+  
geom_histogram(aes(y=..density..),binwidth=1,  
colour="black",fill="white")+  
geom_density(alpha=.1,fill="#00FFFF")
```

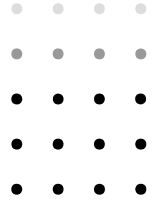
### Find More:

<http://www.sthda.com/english/wiki/ggplot2-histogram-plot-quick-start-guide-r-software-and-data-visualization>





# VISUALIZATION

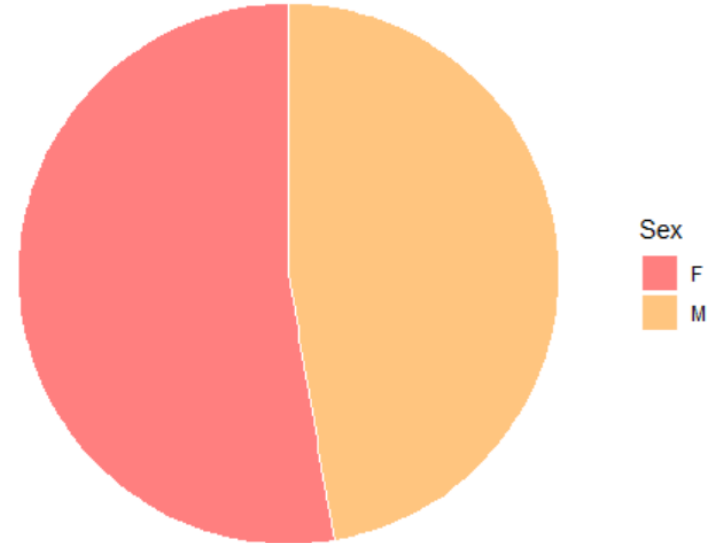


## Pie Chart

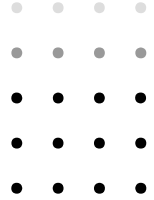
```
library(tidyverse)
sex.c=count(data,Sex)
sex.c=sex.c %>% mutate(perc=round(n*100/sum(n),1))
sex.c=sex.c %>%
  arrange(desc(Sex)) %>%
  mutate(y_pos = cumsum(perc)-0.5*perc)
sex.c %>%
  ggplot(aes(x="",perc, fill=Sex)) +
  geom_bar(width=1,stat="identity",color="white",alpha=
.5) +
  coord_polar("y", start=0)+
  scale_fill_manual(values = rainbow(11)) +
  theme_void()
```

### Find More:

<http://www.sthda.com/english/wiki/ggplot2-pie-chart-quick-start-guide-r-software-and-data-visualization>



# VISUALIZATION

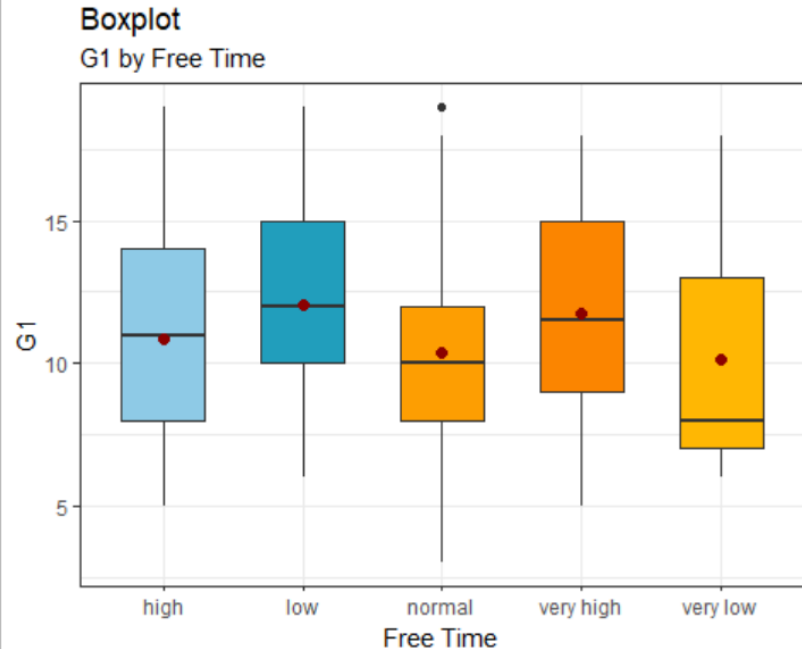


## Boxplot

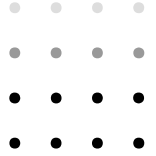
```
ggplot(data,aes(freetime,G1,fill=freetime))+geom_
boxplot(width=0.6)+
stat_summary(fun=mean,geom="point",size=2,col
or="darkred")+theme_bw()+theme(legend.position
="none")+labs(title="Boxplot", subtitle="G1 by Free
Time",x="Free
Time",y="G1")+scale_fill_manual(values=c("#8ECAE
6","#219EBC","#FD9E02","#FB8500","#FFB703"))+
theme(legend.position="none")
```

### Find More:

<http://www.sthda.com/english/wiki/ggplot2-box-plot-quick-start-guide-r-software-and-data-visualization>



# VISUALIZATION

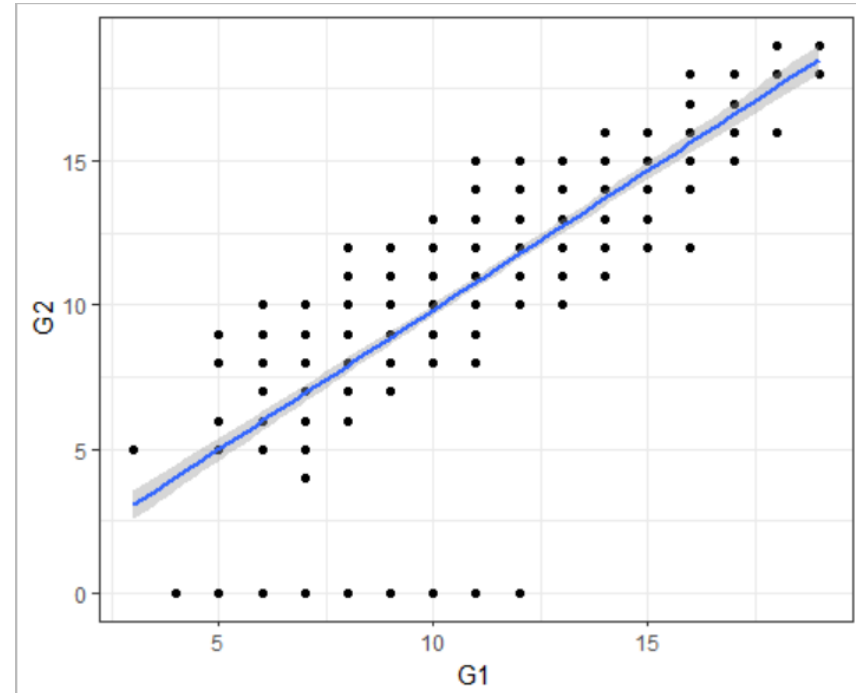


## Scatterplot

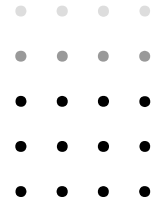
```
ggplot(data, aes(x=G1, y=G2)) +  
  geom_point()+  
  geom_smooth(method=lm)+theme_bw()
```

### Find More:

<http://www.sthda.com/english/wiki/ggplot2-scatter-plots-quick-start-guide-r-software-and-data-visualization>



# DESCRIPTIVE STATISTICS



- Mean

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- Variance

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

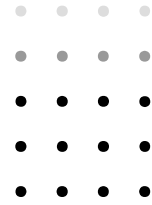
- Covariance

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

- Correlation

$$\text{cor}(x, y) = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$

# MEAN

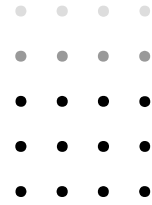


- Formula

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- Algorithm

1. Add all value of  $x$  ( $\sum_{i=1}^n x_i$ )
2. Divide the output of step 1 with  $n$



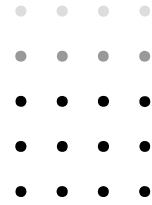
# VARIANCE

- Formula

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

- Algorithm

1. Calculate each  $x_i - \bar{x}$
2. Add all value of  $(x_i - \bar{x})^2$
3. Divide the output of step 2 with  $n - 1$



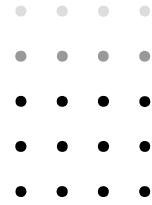
# COVARIANCE

- Formula

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

- Algorithm

1. Calculate each  $x_i - \bar{x}$  and  $y_i - \bar{y}$
2. Multiply  $x_i - \bar{x}$  and  $y_i - \bar{y}$
3. Add all value of  $(x_i - \bar{x})(y_i - \bar{y})$
4. Divide the output of step 3 with  $n - 1$



# CORRELATION

- Formula

$$\text{cor}(x, y) = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$





# THANKS

---

<https://intip.in/KomstatC2023>

