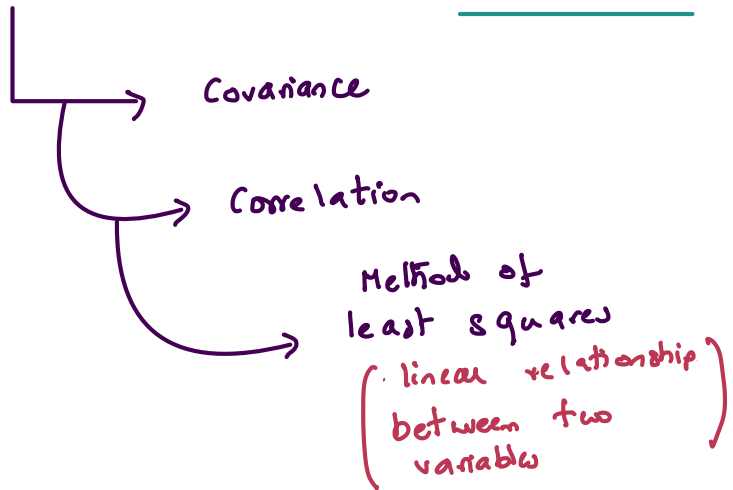


## Bivariate Data

## Agenda for week 9



- 
- Exploratory Data Analysis
  - Case study ÷ Maternal smoke vs infant health
  - Reference for study

# Bivariate Data

- Many Data consists of 2 variables.
  - One of them is a **Dependent variable**, sometimes referred to as **Response variable**
  - One of them being a **independent variable**, sometime referred to as **Predictor or Explanatory variable**.
- There are situations when there is one response variable and multiple explanatory variables. We will not discuss them in this course.
- We will focus on Bivariate Data.

# Examples

- Maternal Smoking and its effect on Birth Weight.
- Attendance in Classes and its effect on Scores in an Exam.
- Age and Heart rate
- Effect of Vitamin C on Toothgrowth

# Data

```
> df = read.table("baby.csv", header = T)
```

```
> head(df)
```

	bwt	smoke
1	120	0
2	113	0
3	128	1
4	123	0
5	108	1
6	136	0

```
> unique(df$smoke)
```

```
[1] 0 1 9
```

```
> df$number= 1:1236
```

- Maternal Smoking vs  
Infant Health

baby.csv

1236 infant

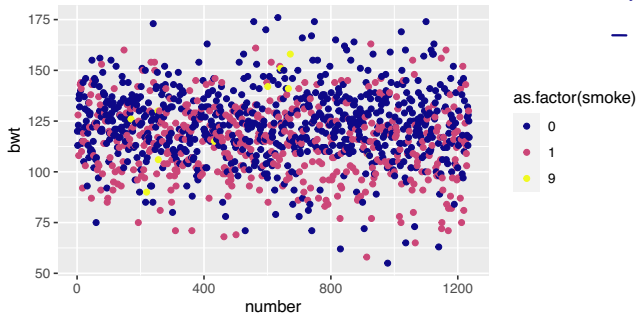
bwt  $\equiv$  birth weight in  
ounces

smoke  $\equiv$   $\begin{cases} 0 - \text{did not smoke} \\ 1 - \text{smoker} \\ 9 - \text{Don't know} \end{cases}$

# Data

Ex: -

- change axis labels
- add title
- change legend

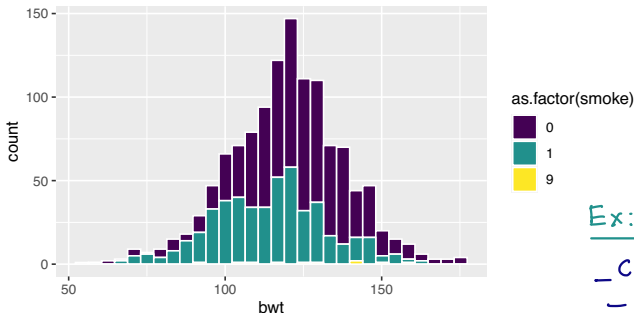


- Scatter plot of birth weight  
layered - smoke as factor.

# Data

Histogram of Birth weight  
- fill it with smoke as  
factor

-(stacked)

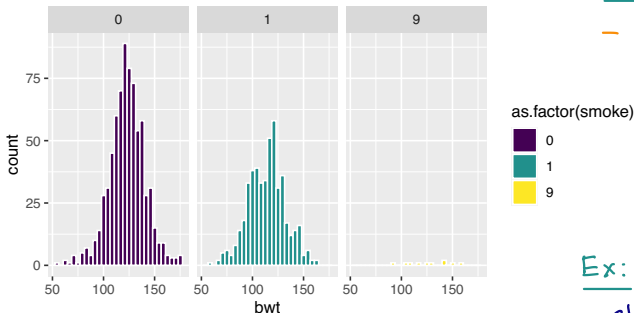


Ex: -

- change axis labels
- add title
- change legend

# Data

Histogram - with Facet - wrap



Ex:-

- plot histogram proportions

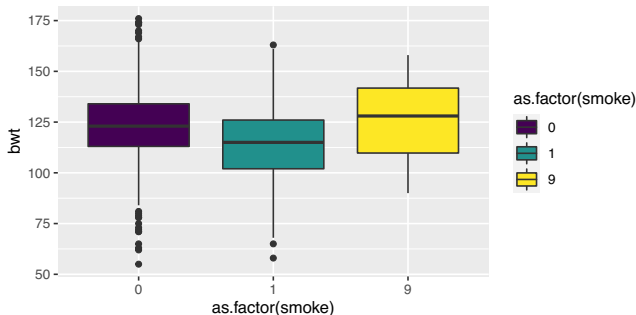
Ex:-

- change axis labels
- add title
- change legend

# Data

- Mean Birth weight  
Did not smoke
- >
- Mean Birth weight  
Did not smoke

Box-plot

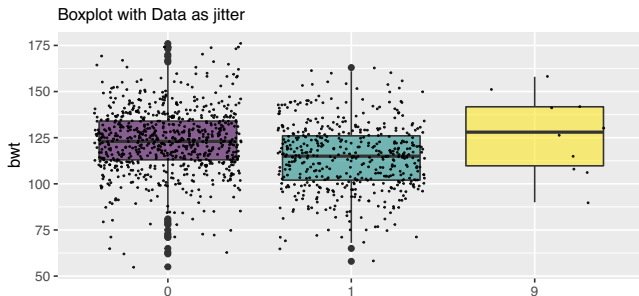


Can this hypothesis be verified?



# Data

Box-plot  
with Data



- How does one analyse the data further?
  - Frame an hypothesis
  - Execute a test on data

# Maternal Smoking and Infant deaths

*Smoking by pregnant women may result in fetal injury, premature birth, and low birth weight.*

- Is this warning to be taken seriously ?
- Dataset: CHDS - Berkeley, California. — (1960-67)
- Taken entirely from Chapter 1 of the book  
*Stat Labs: Mathematical Statistics Through Applications*  
by Deborah Nolan and Terry P. Speed
- Local Copy: Chapter 1 File
- Check website:  
<https://www.stat.berkeley.edu/users/statlabs/>

# Maternal Smoking and Infant deaths

- At birth, measurements on the baby were recorded. They included the baby's length, weight, and head circumference.
- `Babies1.data` is a subset of this information collected. Contains data on:
  - weight of 1236 baby boys born during one year of the study who lived at least 28 days and who were single births (i.e., not one of a twin or triplet).
  - if the mother smoked during her pregnancy.

# Maternal Smoking and Infant deaths

- Epidemiological Studies indicate that birth weight is a measure of the baby's maturity (health).
- Reading:- From Chapter 1, please read sections on *Fetal Development*, *Rubella* and *Physical Model*.

# Maternal Smoking and Infant deaths

- Is there a difference between birth weights of babies born to smokers and those born to non-smokers ?
- Is the difference important to the health and development of the baby ?

# Maternal Smoking and Infant deaths

[Yer71] Yerushalmy:

- The babies were grouped according to their birth weight;
- Within each group, the numbers of babies that died in the first 28 days after birth for smokers and nonsmokers were compared.
- To accommodate the different numbers of babies in the groups, rates instead of counts are used in making the comparisons.

# Maternal Smoking and Infant deaths

He calculated Neonatal mortality rates per 1000 births by birth weight (grams) for live-born infants of white mothers, according to smoking status. They are given in table

Not adjusted

- Age  
- other  
health  
fact

Weight in Grams	Nonsmoker	Smoker
$\leq 1500$	792	562
1500-2000	406	346
2000-2500	78	27
2500-3000	11.6	6.1
3000-3500	2.2	4.5
3500+	3.8	3.6

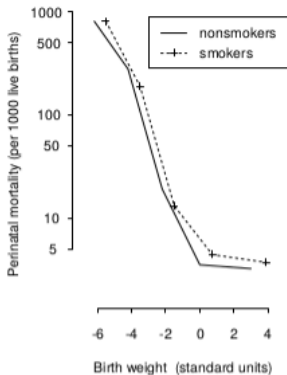
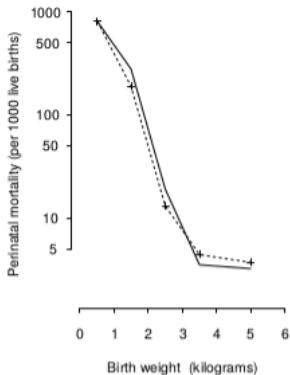
# Maternal Smoking and Infant deaths

[Yer71] Yerushalmy found that although low birth weight is associated with an increase in the number of babies who die shortly after birth, the babies of smokers tended to have much lower death rates than the babies of nonsmokers.



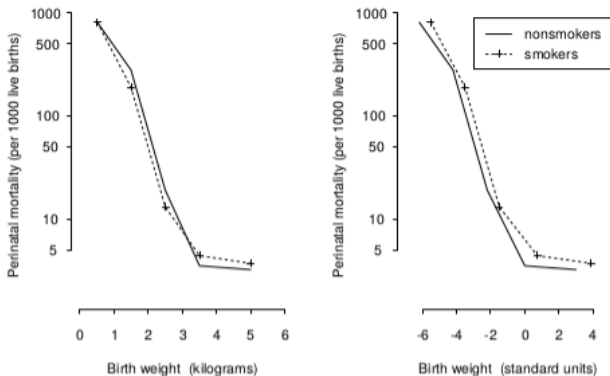
# Maternal Smoking and Infant deaths

Wilcox and Russell [WR86] advocate grouping babies according to their relative birth weights. Plotted Mortality rate for perinatal stage.



# Maternal Smoking and Infant deaths: Conclusion 2

[Wr86] They found that for babies born at term, smokers have higher rates of perinatal mortality in every standard unit.



# Maternal Smoking and Infant deaths: Two conclusions?

- [Y71]
  - not adjusted for the mother's age.
  - young smoker versus old non-smoker. Medical factors.
- [MKLS88] Adjusted for above:
  - found that the risk of neonatal death for babies who were born at 32+ weeks gestation is roughly the same for smokers and nonsmokers.
  - It was also found that the smokers had a higher rate of very premature deliveries (20-32 weeks gestation), and so a higher rate of early fetal death.

# Maternal Smoking and Infant deaths

[Wr86]

- Babies born to smokers tend to be smaller, the mortality curve is shifted to the right relative to the nonsmokers curve. [verify?](#)
- If the babies born to smokers are smaller but otherwise as healthy as babies born to nonsmokers, then the two curves in standard units should roughly coincide.