

# ggplot-Regression

## Regression with ggplot

1. `geom_smooth`
2. `se=FALSE`
3. `lm(Y~X)`

## Data

```
weight <- c(6.2,7.2,7.6,8.3,8.6,9.2,9.5,10.0,10.4,10.6,11.1,
            6.9,7.6,8.4,8.8,9.3,9.6,10.1,10.5,10.8,11.4,11.7)
age <- c(4,6,8,10,12,14,16,18,20,22,24,
        4,6,8,10,12,14,16,18,20,22,24)
sex <- c('F','F','F','F','F','F','F','F','F','F','F',
        'M','M','M','M','M','M','M','M','M','M','M')
df <- data.frame(weight,age,sex)
str(df)
```

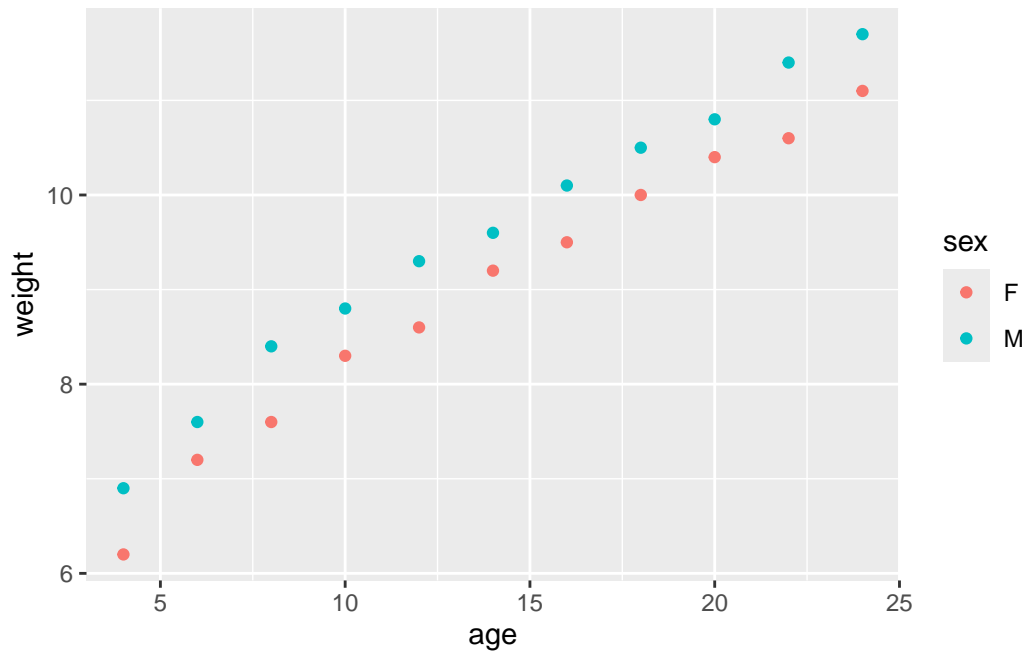
```
'data.frame':  22 obs. of  3 variables:
 $ weight: num  6.2 7.2 7.6 8.3 8.6 9.2 9.5 10 10.4 10.6 ...
 $ age   : num   4 6 8 10 12 14 16 18 20 22 ...
 $ sex   : chr  "F" "F" "F" "F" ...
```

---

---

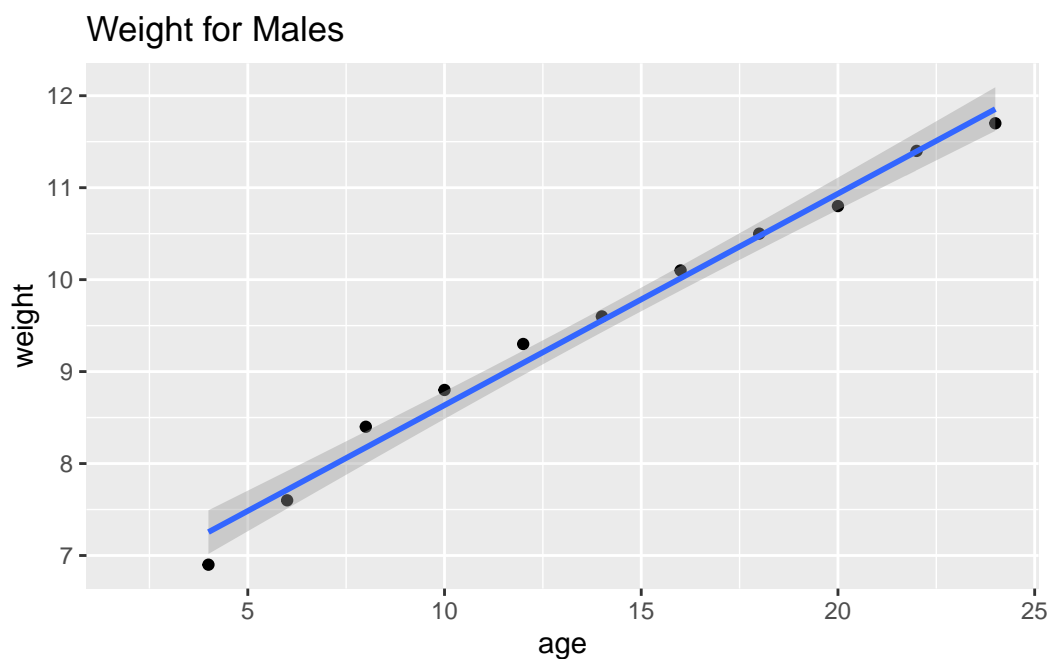
## Sorted Scatterplot by sex

```
df1 <- arrange(df, -desc(age))
ggplot(df1, aes(x=age, y=weight, col=sex))+
  geom_point()
```



## Regression on male data

```
df2 <- df1 %>%
  filter(sex=='M')
ggplot(df2, aes(x=age, y=weight))+
  xlim(2, 24)+
  geom_point()+
  geom_smooth(method="lm")+
  labs(
    title="Weight for Males"
  )
```



```
model <- lm(weight~age, data=df2)
summary(model)
```

Call:

```
lm(formula = weight ~ age, data = df2)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.35455	-0.12455	0.02545	0.12545	0.22545

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	6.334545	0.136201	46.51	4.92e-12 ***
age	0.230000	0.008866	25.94	9.06e-10 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.186 on 9 degrees of freedom

Multiple R-squared: 0.9868, Adjusted R-squared: 0.9853

F-statistic: 673 on 1 and 9 DF, p-value: 9.062e-10