

# Bài giảng 8a: Giới thiệu ggplot2

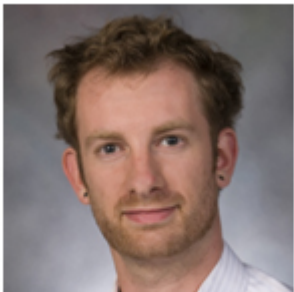
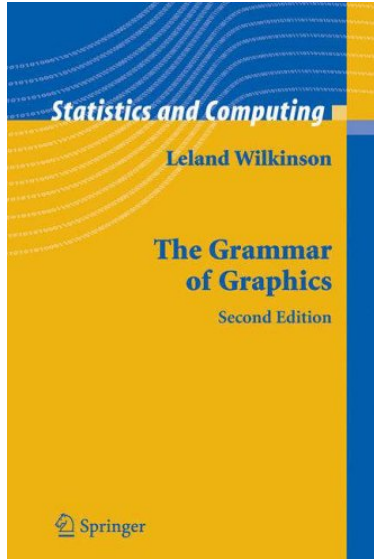
**Nguyễn Văn Tuấn**

Garvan Institute of Medical Research, Australia  
Ton Duc Thang University, Vietnam

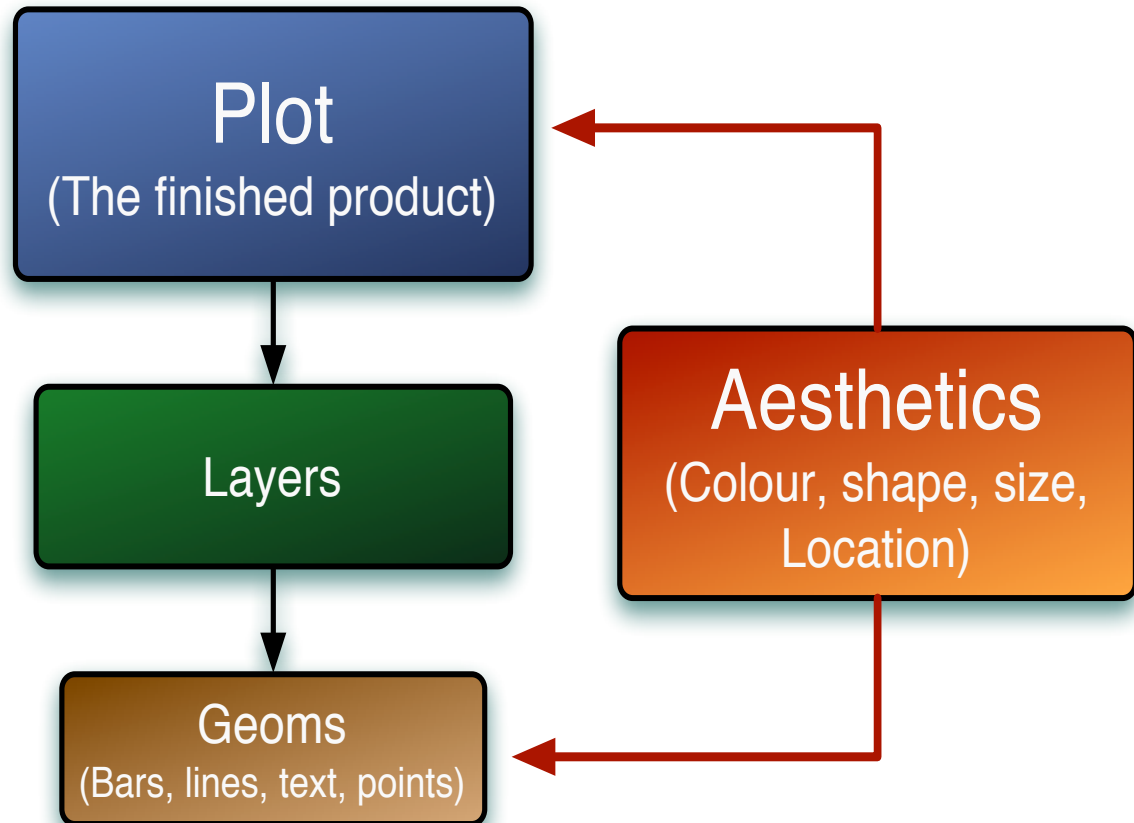
# Nội dung

- Giới thiệu "mô hình" ggplot2
- Dữ liệu thực tế
- Vẽ biểu đồ phân bố với hàm ggplot

# Mô hình ggplot2



Hadley Wickham



# Biểu đồ có nhiều **layer** (lớp)

- **Layer 1:** Định nghĩa biến cần phân tích (vẽ): **aes()**
- **Layer 2:** Thể loại biểu đồ: **geom\_xxx()**
- **Layer 3:** Màu, kích thước, v.v.: **theme\_xxx()**
- **v.v.**

# Aesthetics (aes)

- Data
- Color
- Size
- Shape
- v.v.

```
aes (col="gender")
```

```
aes (shape="group")
```

```
aes (x = mpg, y = wt)
```

# Geometric objects (geom)

Geometric object	Ý nghĩa
geom_ <b>histogram</b> (x=)	Biểu đồ phân bố
geom_ <b>boxplot</b> (x=)	Biểu đồ hộp
geom_ <b>bar</b> (x=)	Biểu đồ thanh
geom_ <b>point</b> (x=, y=)	Biểu đồ điểm
geom_ <b>line</b> (x=, y=)	Biểu đồ đường thẳng
geom_ <b>smooth</b> (x=, y=)	Biểu đồ smooth
geom_ <b>density</b> (x=, y=)	Biểu đồ với xác suất phân bố
geom_ <b>errorbar</b> (x=)	Biểu đồ với sai số chuẩn

# Hàm chính: ggplot

```
p = ggplot(data, aes(x, y))
```

```
# thêm layers
```

```
p + geom_xxx() + theme_xxx()
```

# Một số lựa chọn phổ biến: ggplot

# thêm labels

```
p + xlab("xx") + ylab("xx") + ggtitle("xx")
```

# thêm limits

```
p + xlim(0,100) + ylim(0,10)
```

# background trắng đen

```
p + theme_bw()
```

```
p + theme_classic()
```

```
p + theme_minimal()
```

# vị trí legend

```
p + theme(legend.position="top/bottom")
```



**Ví dụ về "vận hành" của  
ggplot2**

# Đọc dữ liệu

```
ob = read.csv("obesity data.csv", header=T,  
na.strings=".")
```

```
attach(ob)
```

```
head(ob)
```

```
> head(ob)
```

	<b>id</b>	<b>gender</b>	<b>height</b>	<b>weight</b>	<b>bmi</b>	<b>age</b>	<b>WBBMC</b>	<b>wbbmd</b>	<b>fat</b>	<b>lean</b>	<b>pcfat</b>
<b>1</b>	<b>1</b>	<b>F</b>	<b>150</b>	<b>49</b>	<b>21.8</b>	<b>53</b>	<b>1312</b>	<b>0.88</b>	<b>17802</b>	<b>28600</b>	<b>37.3</b>
<b>2</b>	<b>2</b>	<b>M</b>	<b>165</b>	<b>52</b>	<b>19.1</b>	<b>65</b>	<b>1309</b>	<b>0.84</b>	<b>8381</b>	<b>40229</b>	<b>16.8</b>
<b>3</b>	<b>3</b>	<b>F</b>	<b>157</b>	<b>57</b>	<b>23.1</b>	<b>64</b>	<b>1230</b>	<b>0.84</b>	<b>19221</b>	<b>36057</b>	<b>34.0</b>
<b>4</b>	<b>4</b>	<b>F</b>	<b>156</b>	<b>53</b>	<b>21.8</b>	<b>56</b>	<b>1171</b>	<b>0.80</b>	<b>17472</b>	<b>33094</b>	<b>33.8</b>
<b>5</b>	<b>5</b>	<b>M</b>	<b>160</b>	<b>51</b>	<b>19.9</b>	<b>54</b>	<b>1681</b>	<b>0.98</b>	<b>7336</b>	<b>40621</b>	<b>14.8</b>
<b>6</b>	<b>6</b>	<b>F</b>	<b>153</b>	<b>47</b>	<b>20.1</b>	<b>52</b>	<b>1358</b>	<b>0.91</b>	<b>14904</b>	<b>30068</b>	<b>32.2</b>

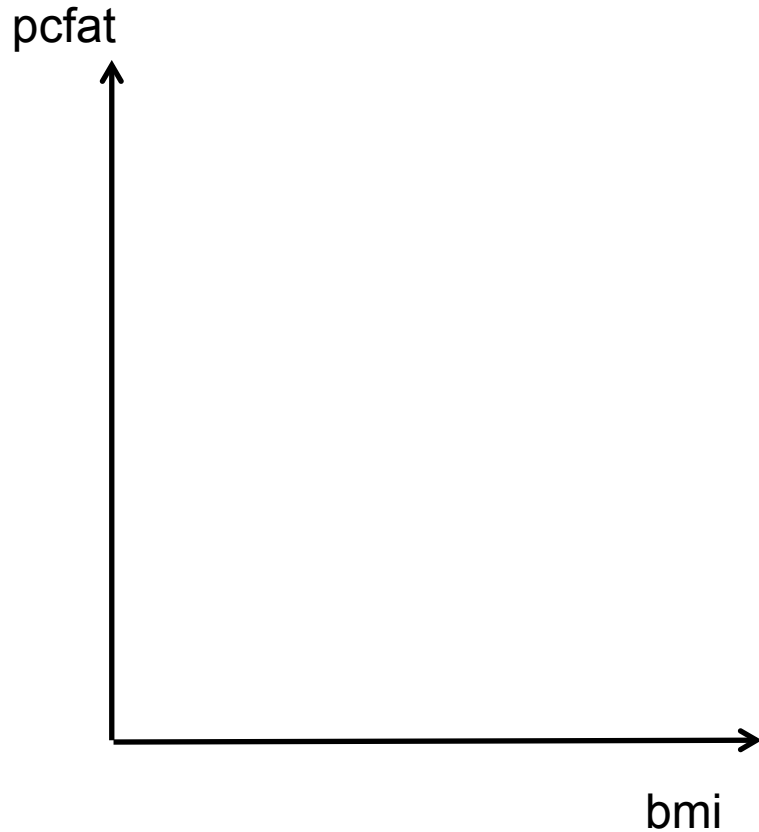
# Dữ liệu

- id height weight bmi age
- gender : giới tính (M, F)
- WBBMC: whole body bone mineral content (g)
- wbbmd: whole body bone mineral density (g/cm<sup>2</sup>)
- fat : fat mass (g)
- lean : lean mass (g)
- pcfat : percent body fat (%)

# Mục đích

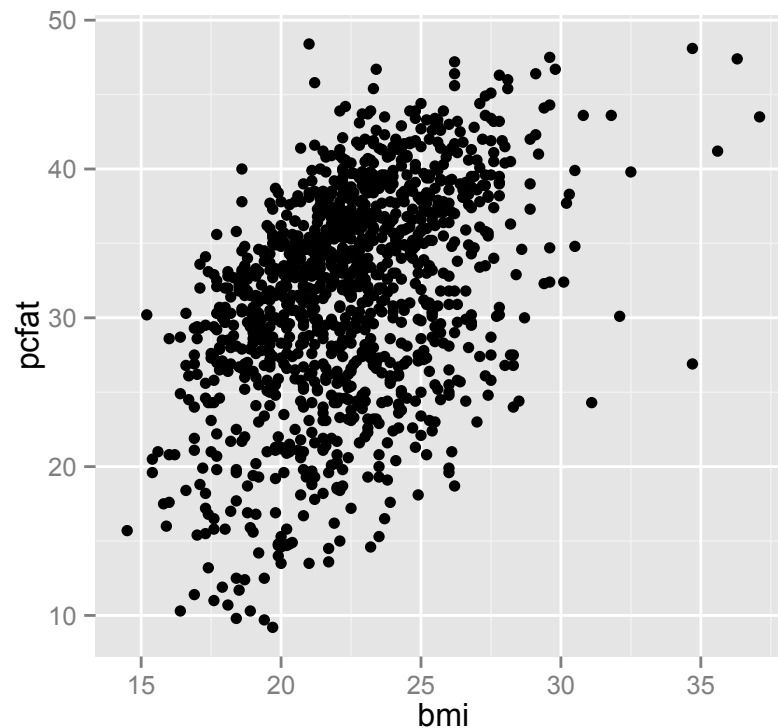
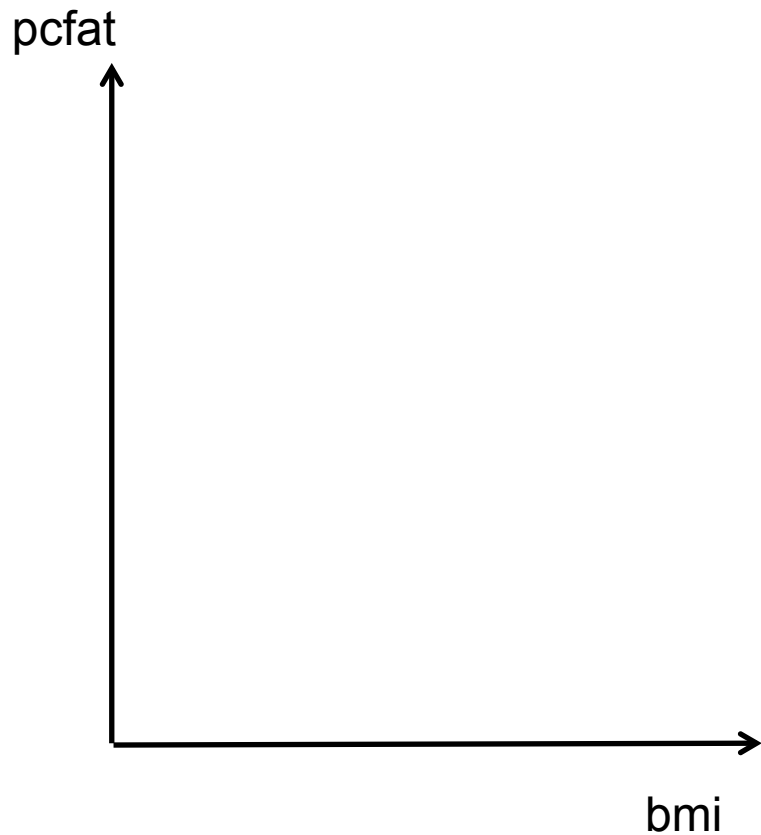
- Muốn mô tả mối liên quan giữa **pcfat** và **bmi**
- Dùng hàm ggplot

# Layer 1: định nghĩa biến cần vẽ



```
p = ggplot(ob, aes(x=bmi,  
y=pcfat))
```

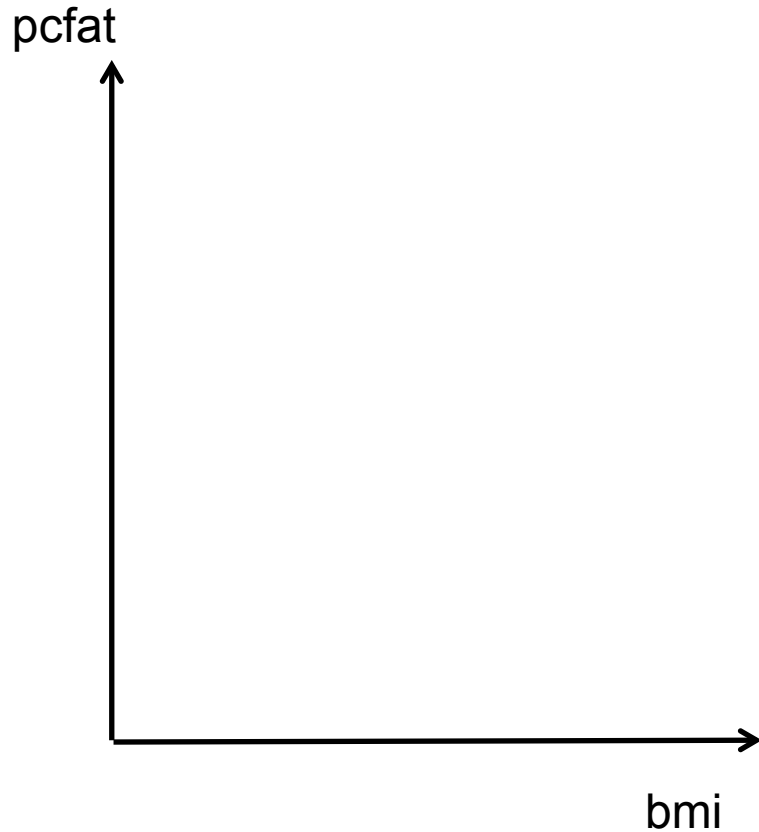
# Layer 2: thể loại biểu đồ



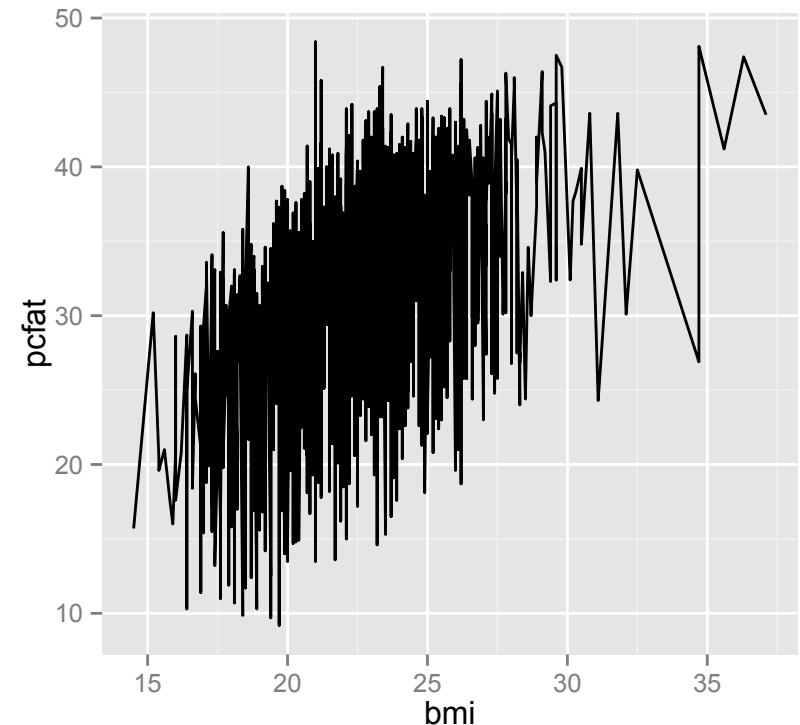
```
p = ggplot(ob, aes(x=bmi,  
y=pcfat))
```

```
p + geom_point()
```

# Layer 2: thể loại biểu đồ

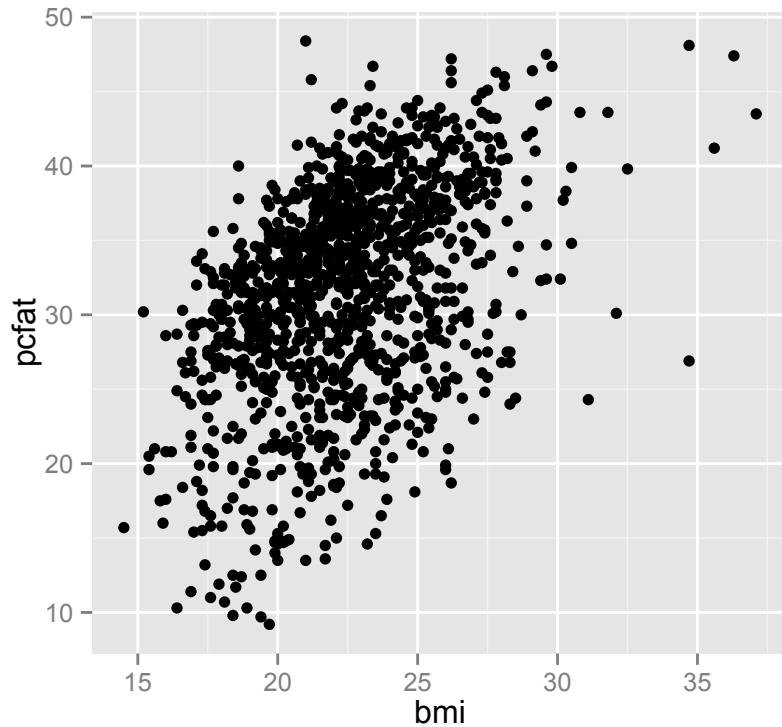


```
p = ggplot(ob, aes(x=bmi,  
y=pcfat))
```

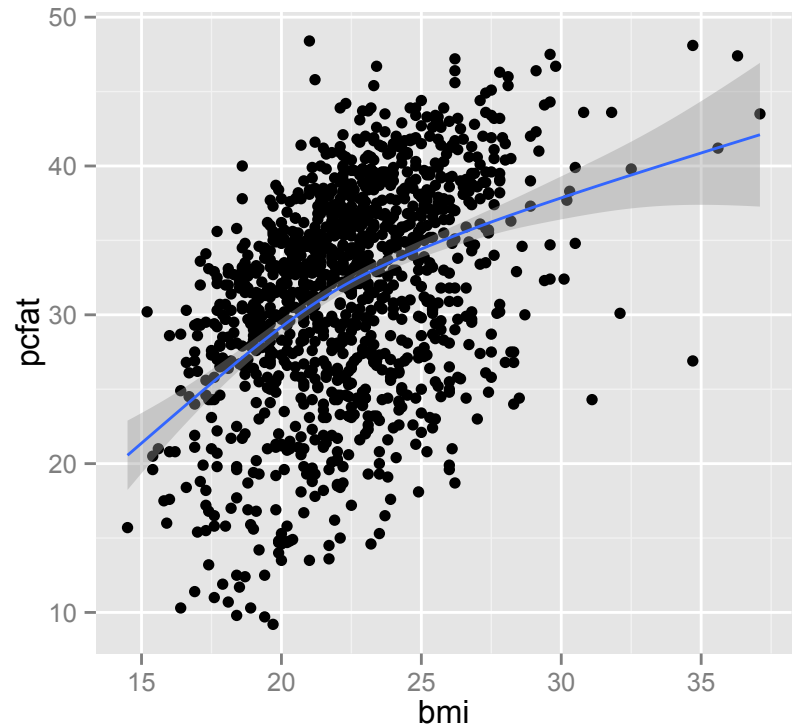


```
p + geom_line()
```

# Layer 3: thêm đường smooth



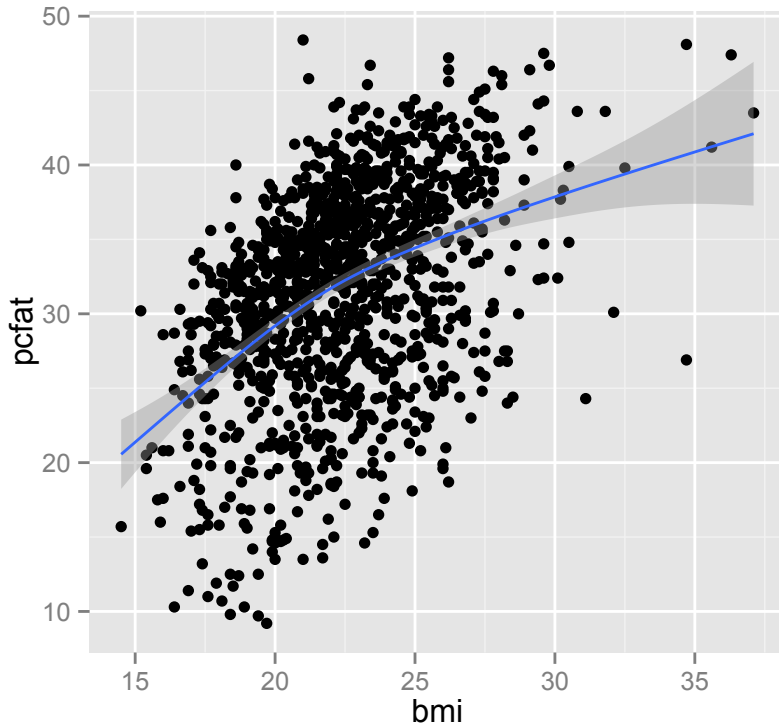
```
p + geom_point()
```



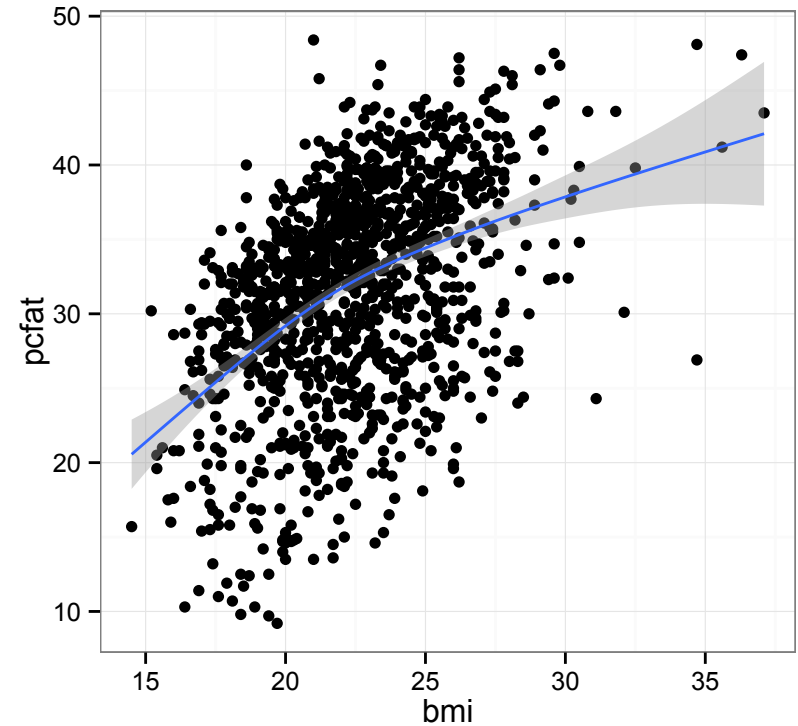
```
p + geom_point() +  
  geom_smooth()
```



# Layer 3: thêm theme

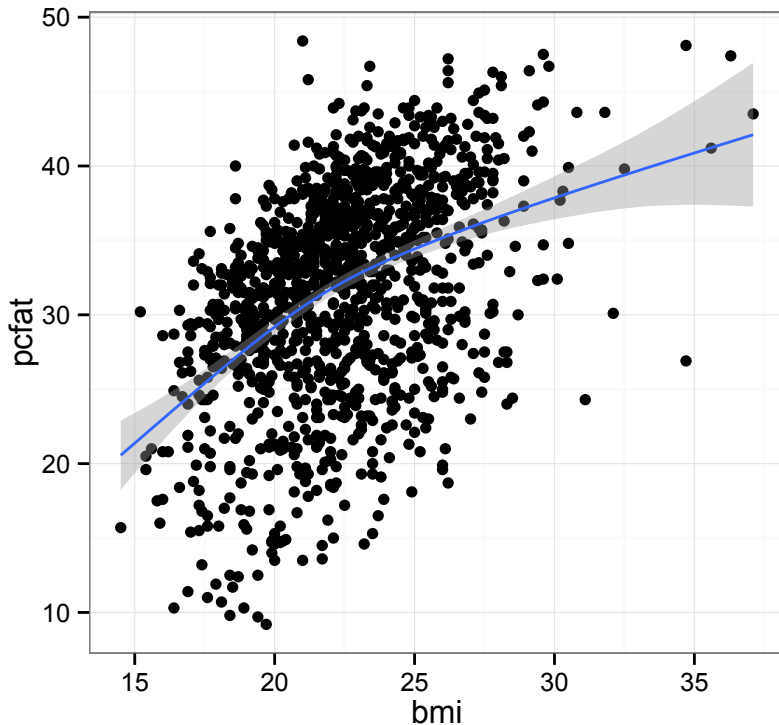


```
p + geom_point() +  
geom_smooth()
```

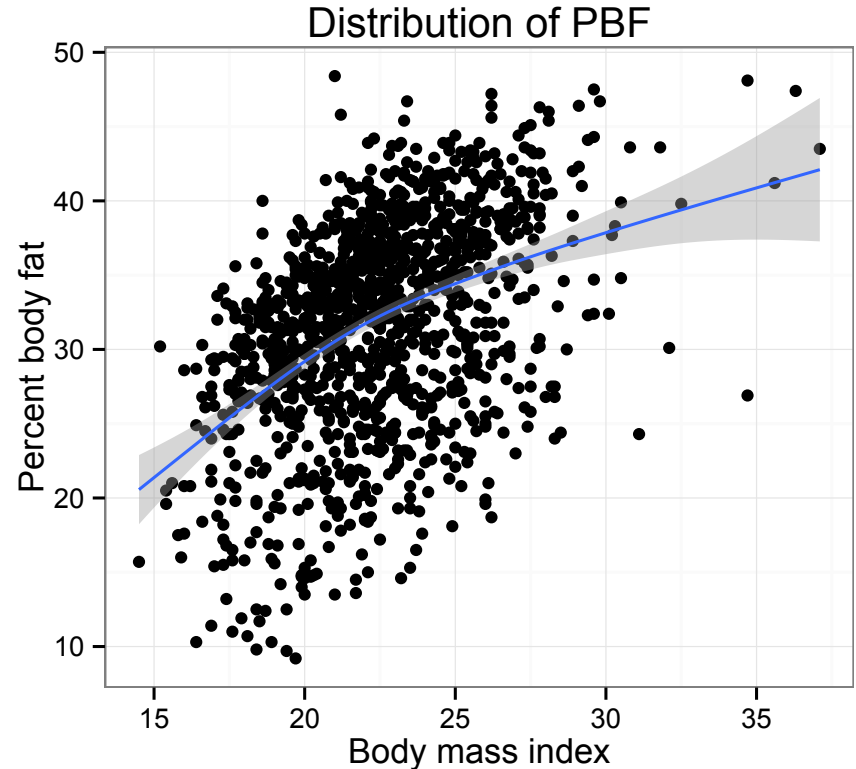


```
p + geom_point() +  
geom_smooth() +  
theme_bw()
```

# Layer 4: thêm labels



```
p + geom_point() +  
geom_smooth() +  
theme_bw()
```

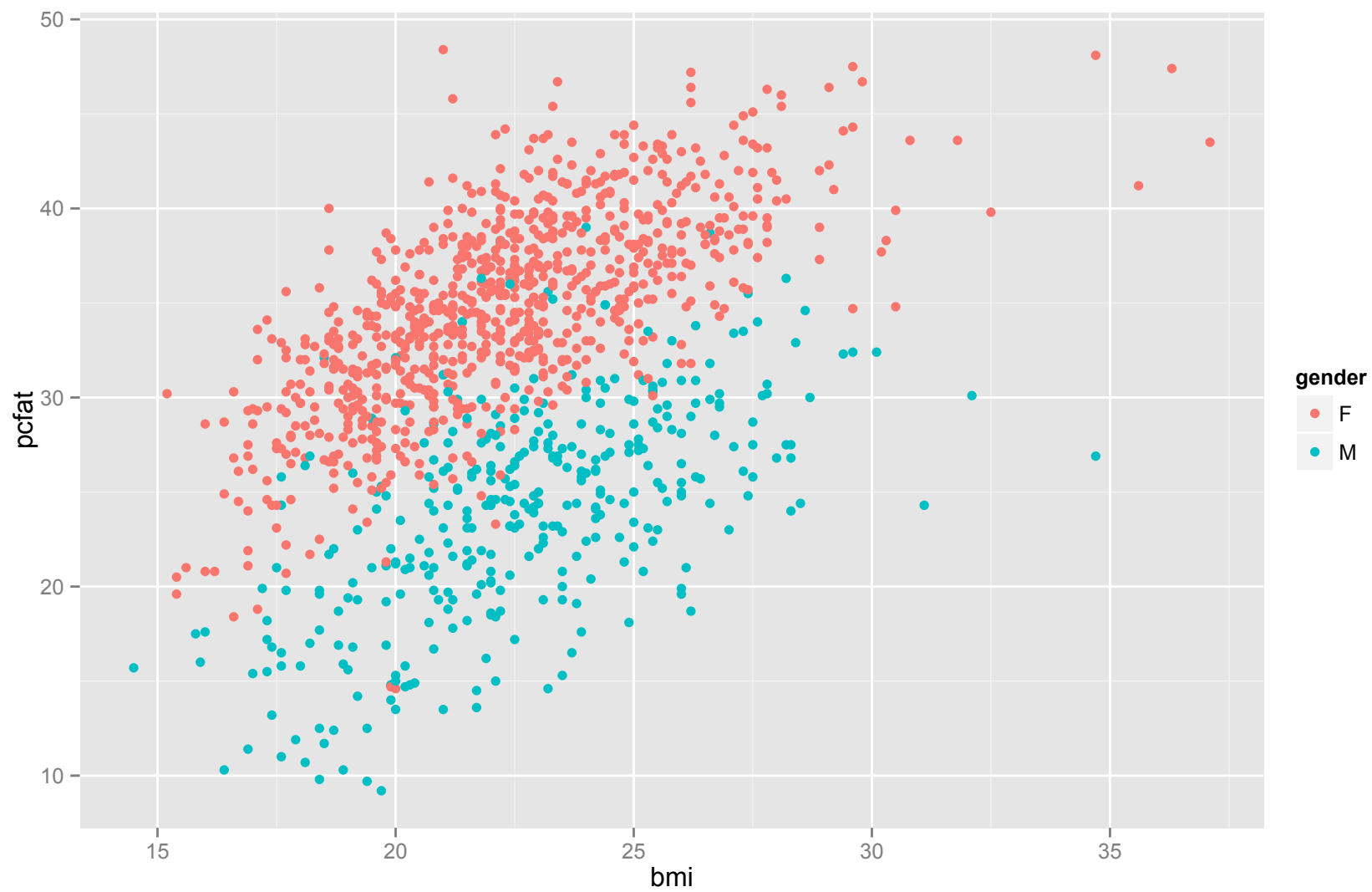


```
p + geom_point() + geom_smooth() +  
theme_bw() + xlab("Body mass  
index") + ylab("Percent body fat")  
+ ggtitle("Distribution of PBF")
```

# Lay 5: thêm nhóm (giới tính)

```
p = ggplot(ob, aes(x=bmi, y=pcfat,  
fill=gender))
```

```
p + geom_point(aes(col=gender))  
p
```



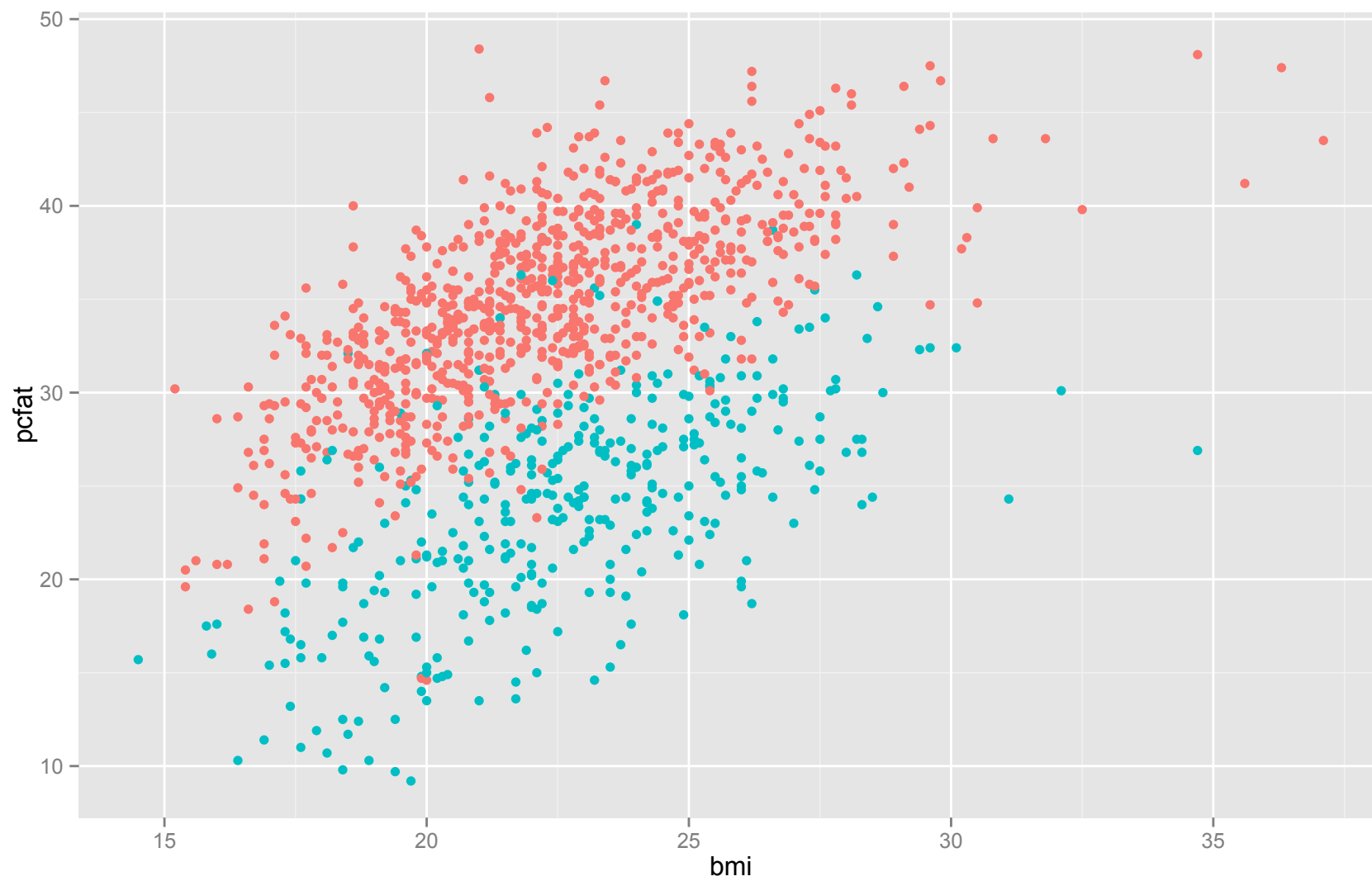
```
p = ggplot(ob, aes(x=bmi, y=pcfat,  
fill=gender))
```

```
p = p + geom_point(aes(col=gender))
```

```
# bô legend
```

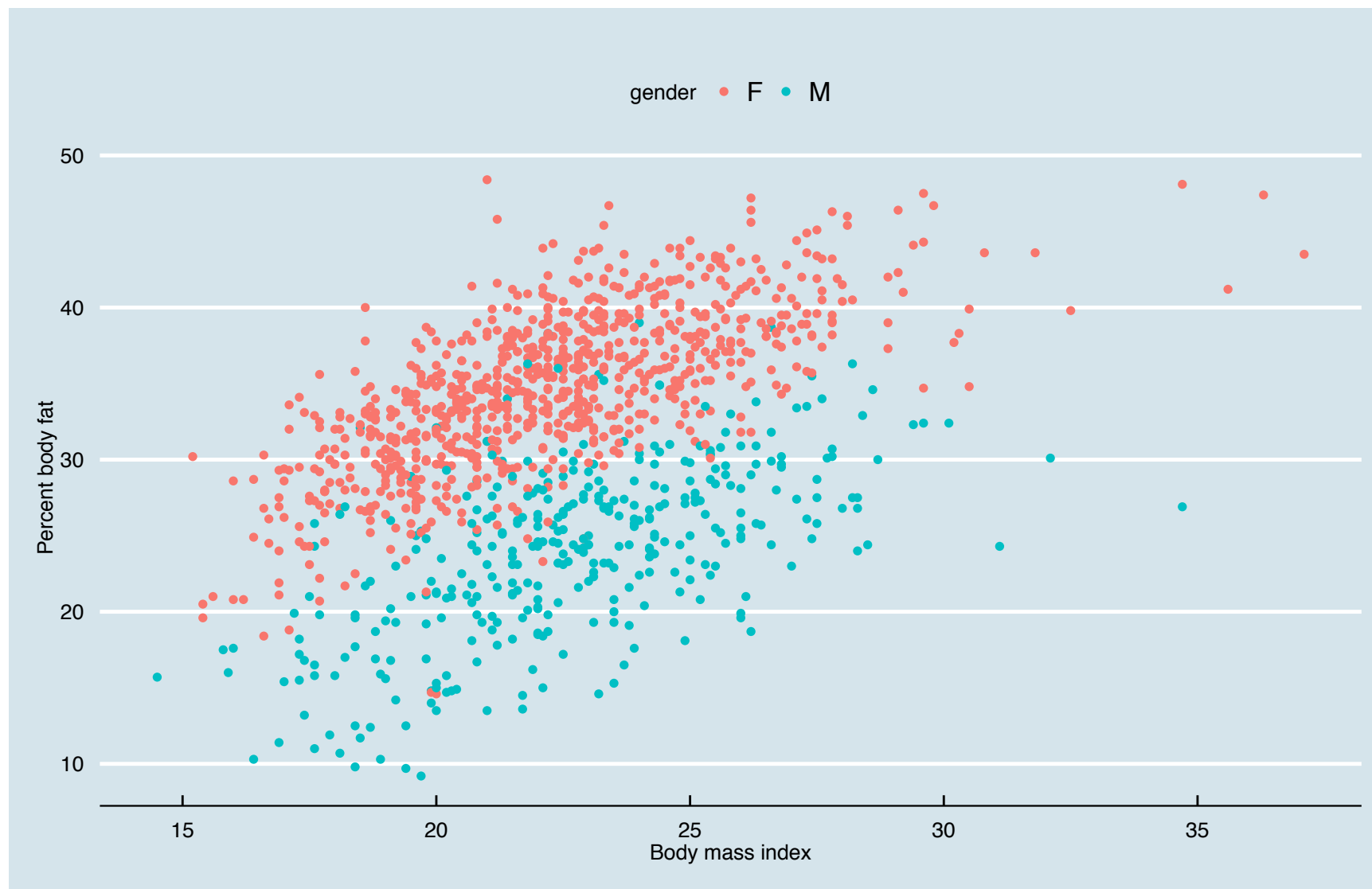
```
p = p + theme(legend.position="none")
```

```
p
```



```
p = ggplot(ob, aes(x=bmi, y=pcfat, fill=gender))
p = p + geom_point(aes(col=gender))

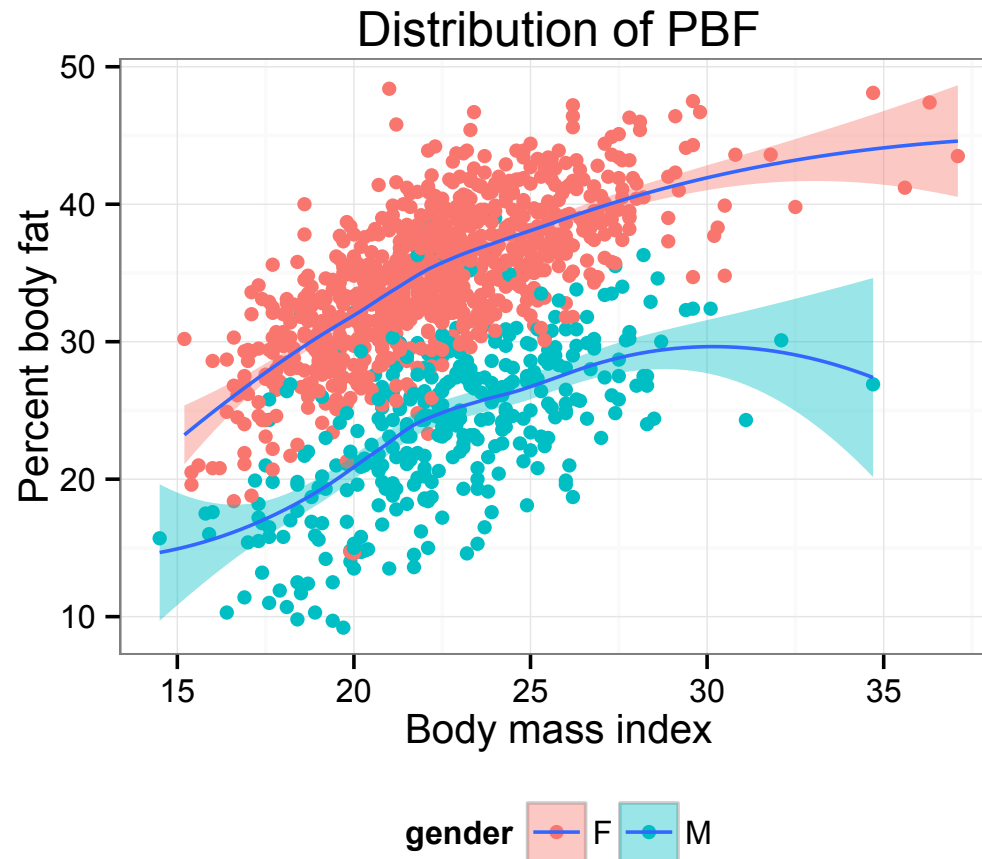
# bỏ legend + thay màu trục hoành và trục tung
p = p + theme(legend.position="none",
axis.text.x=element_text(angle=30, color="black",
size=10),
axis.text.y=element_text(color="black"))
# thêm label
p = p + xlab("Body mass index") + ylab("Percent
body fat")
#theme economist
require(ggthemes)
p = p + theme_economist()
```





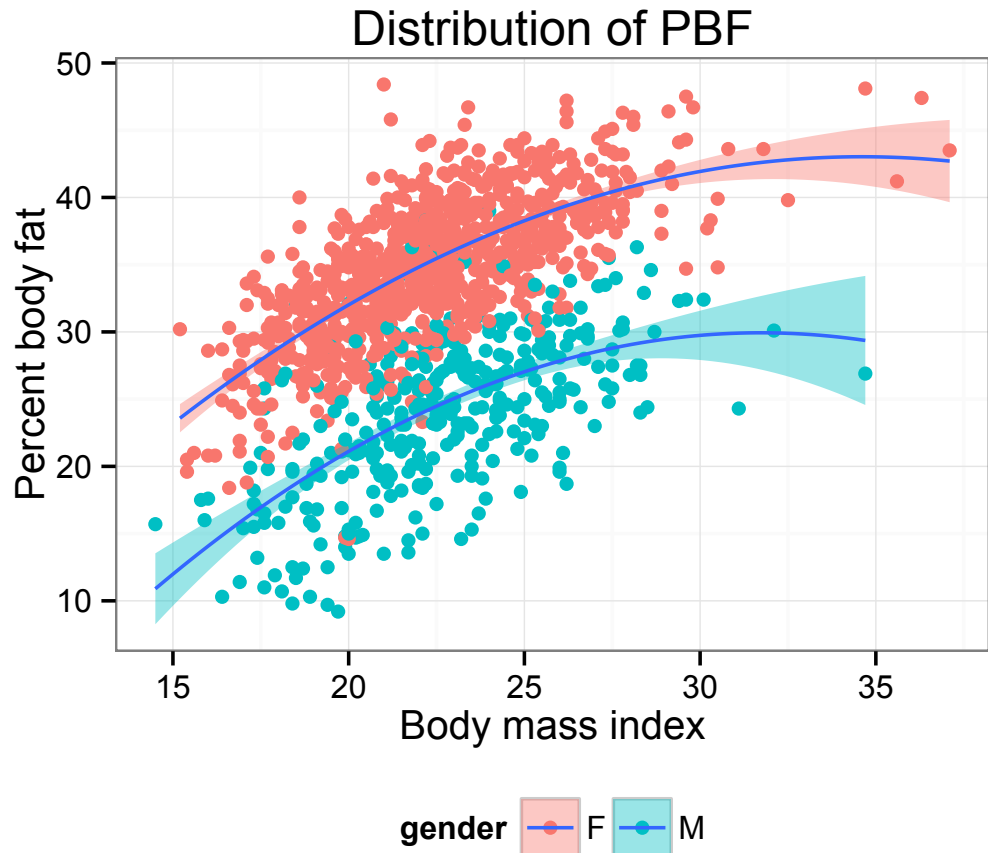
```
p = ggplot(ob, aes(x=bmi, y=pcfat, fill=gender))
```

```
p + geom_point(aes(col=gender)) + geom_smooth() +  
theme_bw() + xlab("Body mass index") + ylab("Percent  
body fat") + ggtitle("Distribution of PBF") +  
theme(legend.position="bottom")
```



```
p = ggplot(ob, aes(x=bmi, y=pcfat, fill=gender))
```

```
p + geom_point(aes(col=gender)) +  
geom_smooth(method="lm", formula=y~x+I(x^2)) +  
theme_bw() + xlab("Body mass index") + ylab("Percent  
body fat") + ggtitle("Distribution of PBF") +  
theme(legend.position="bottom")
```



# Tóm lược

- ggplot2 – một phát triển quan trọng của R
- Cung cấp phương tiện rất tốt cho các biểu đồ chất lượng cao
- Mô hình ggplot2 là nhiều layers: `aes()`, `geom_xxx()`, và `theme_xxx()`