

Bài giảng 7a: Biểu đồ với R

Nguyễn Văn Tuấn

Garvan Institute of Medical Research, Australia

Đại học Tôn Đức Thắng, Việt Nam

Số liệu, số liệu, số liệu

Garvanid	age	gender	actn3	weight	height	fnbmd	lsbmd	trbmd	wtbmd	lean	fat	quadstreng th
8	67.4	Female	RX		72	166	0.97	1.33	0.85	0.85	.	18
9	68.5	Male	XX		87	184	1.01	1.49	0.87	0.84	.	36
10	62.1	Female	RR		72	173	0.84	1.21	0.62	0.71	.	14
27	64.6	Female	RX		85	167	0.86	1.07	0.78	0.68	.	27
28	76.2	Female	RX		48	153	0.65	0.87	0.5	0.46	.	20
33	74.1	Female	RX		52	156	0.83	0.85	0.71	0.61	32.61	22
34	75.3	Female	RR		70	160	0.79	1.19	0.7	0.51	.	8
36	62.2	Male	XX		97	171	1.16	1.44	1.16	0.94	.	31
37	59.8	Female	XX		60	161	0.79	0.91	0.61	0.63	.	35
38	66.4	Male	RR		64	170	0.85	1.12	0.83	0.71	.	30
40	67.7	Male	RX		82	179	0.75	1.07	0.96	0.44	.	41
41	65.1	Male	RR		101	174	0.94	1.22	0.88	0.65	.	45
42	62.1	Female	XX		82	151	0.99	1.42	0.96	0.94	.	22
47	72.5	Female	RX		56	150	0.62	0.75	0.56	0.42	.	12
49	72.2	Female	RR		46	154	0.58	0.79	0.49	0.38	.	19
51	68.7	Female	RX		65	163	0.79	1.28	0.8	0.66	.	.
53	80.4	Male	RR		76	178	0.99	1.23	0.99	0.78	.	18
56	69.3	Female	RX		55	162	0.75	0.73	0.52	0.58	.	26
57	63	Male	RR		86	160	.	1.49	.	.	.	30
58	60.9	Female	RX		58	161	0.59	0.76	0.5	0.5	.	32
60	58.6	Female	RX		64	151	0.94	0.94	0.73	0.68	.	21
61	59.8	Male	RX		86	184	1.17	1.46	1.1	1.01	.	46
62	59	Female	RR		75	169	0.95	1.05	0.64	0.78	.	9
64	71.5	Female	XX		55	155	0.63	0.94	0.57	0.46	.	26
65	69.3	Male	XX		114	177	1.19	1.38	1.25	0.96	.	46

Chromosome				ensembl_id										
chr	start	end	promoter	genes	id	147_met	163_met	164_met	165_met	173_met	205_met	243_met	244_met	245_met
1	12990440	12990440	NO	Slco5a1	ENSMUSG000000025 938	41.1765	21.4286	26.9231	32.1429	36.3636	22.7273	4.16667	4.16667	0
	23656456	23656456	NO			29.5455	48.7805	37.931	45.4545	33.3333	37.7778	7.69231	11.7647	6.06061
	23656490	23656490	NO			43.1818	51.2195	48.2759	42.8571	35.5556	42.2222	23.0769	21.5686	15.1515
1	36828786	36828786	NO	Tmem131	ENSMUSG000000026 116	40.7407	64.7059	66.6667	50	66.6667	71.4286	100	94.4444	100
	38210579	38210579	NO	Aff3	ENSMUSG000000037 138	51.4286	64.1026	50	52.6316	56	60.6061	3.57143	0	14.2857
	38210636	38210636	NO	Aff3	ENSMUSG000000037 138	66.6667	75	50	57.8947	52	60.6061	17.8571	6.89655	22.8571
1	38210644	38210644	NO	Aff3	ENSMUSG000000037 138	61.1111	80	54.5455	57.8947	56	54.5455	10.7143	3.44828	20
	38210645	38210645	NO	Aff3	ENSMUSG000000037 138	71.4286	69.2308	66.6667	61.5385	63.6364	46.1538	0	0	14.2857
	38830154	38830154	NO	Lonrf2	ENSMUSG000000048 814	51.0638	44.0476	75	68.6567	42.8571	59.6244	93.9759	85.124	79.8479
1	38830168	38830168	NO	Lonrf2	ENSMUSG000000048 814	46.5241	50	63.8889	73.1343	50.4464	64.3193	86.747	92.8375	88.5932
	39012738	39012738	NO			45.8333	50	35.7143	45.6522	49.2537	47.9167	87.8049	90.3614	88
	39012812	39012812	NO			40.8163	46.4286	48.2759	40.8163	48.5294	45.8333	82.9268	78.3133	85.5263
1	39012885	39012885	NO			50	42.8571	55.1724	48.9796	52.9412	46.8085	85	90.3614	84
	41157209	41157209	NO			88.8889	92.8571	80	72.4138	86.1111	82	7.69231	14	28.3333
1	59484325	59484325	NO	Fzd7	ENSMUSG000000041 075	57.5758	50.9434	50	44.8276	31.5789	42.8571	0	3.22581	2.7027
	59484440	59484440	NO	Fzd7	ENSMUSG000000041 075	91.3043	89.1892	68	95.1219	63.1579	81.1321	33.3333	16.129	28.9474
	59484455	59484455	NO	Fzd7	ENSMUSG000000041 075	91.3043	83.7838	68	85.3659	63.1579	84.9057	33.3333	16.129	15.7895
1	59484487	59484487	NO	Fzd7	ENSMUSG000000041 075	100	94.5946	60	90.2439	68.4211	84.9057	22.2222	19.3548	23.6842

Câu hỏi

- Làm sao hình dung dữ liệu bằng hình ảnh?
- Làm sao tìm qui luật trong rừng dữ liệu?
- Làm sao định lượng hóa các mối tương quan?
- V.V.

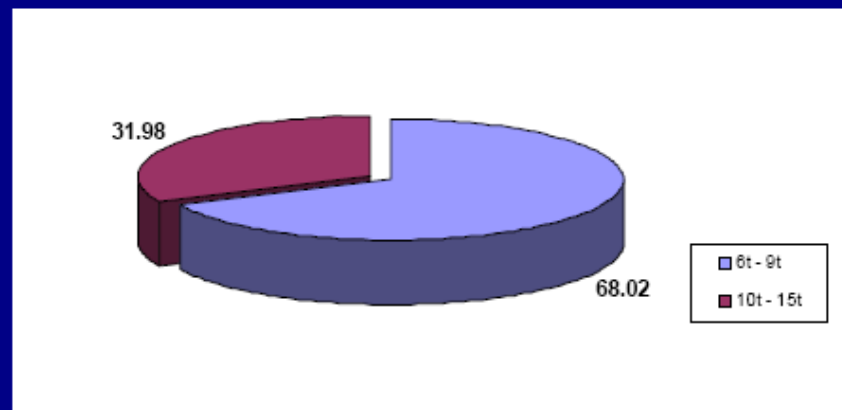
Biểu đồ trong khoa học rất quan trọng

- "*A picture is worth a thousand words*"
- Dễ gây ấn tượng ở người đọc
- Biểu đồ có giá trị lâu dài, có thể đi vào lịch sử
- Đòi hỏi phải suy nghĩ trong thiết kế biểu đồ

Nguyên tắc soạn biểu đồ

- Nói lên sự thật
- Tối đa hoá tỉ lệ dữ liệu trên mực in
- Tối đa hoá mật độ dữ liệu
- Nghiêm chỉnh! (Tránh hoa hòè)

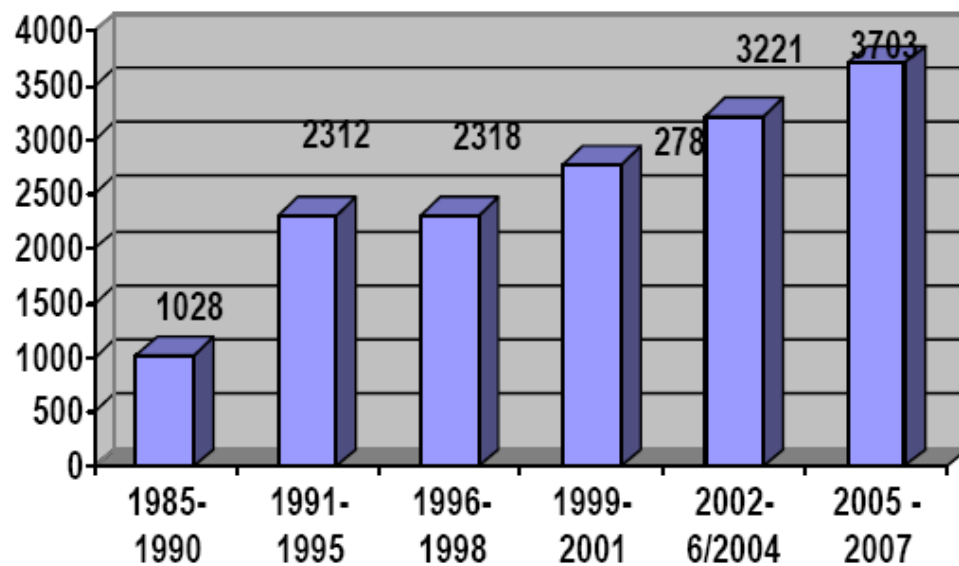
Chart 3.1. Age distribution



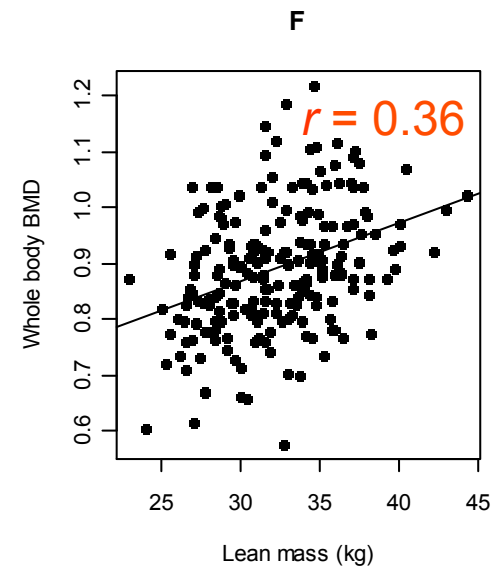
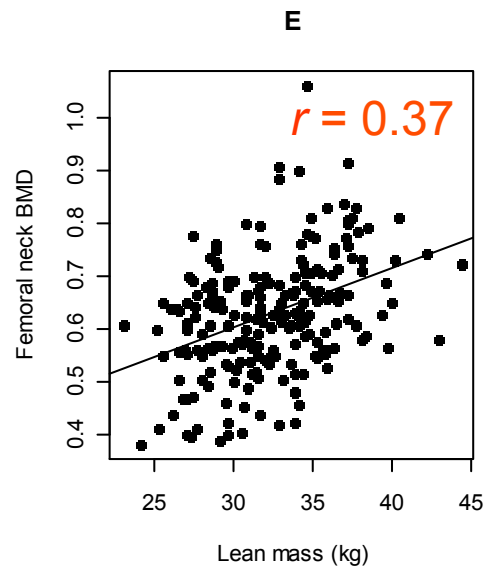
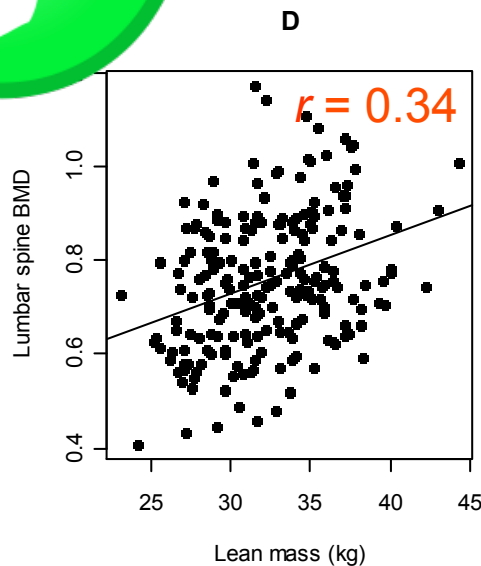
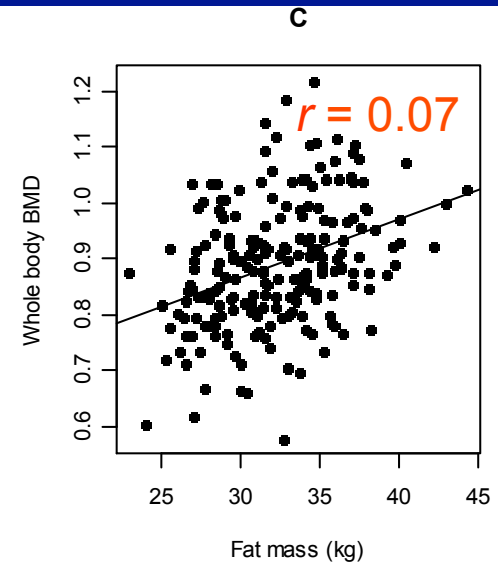
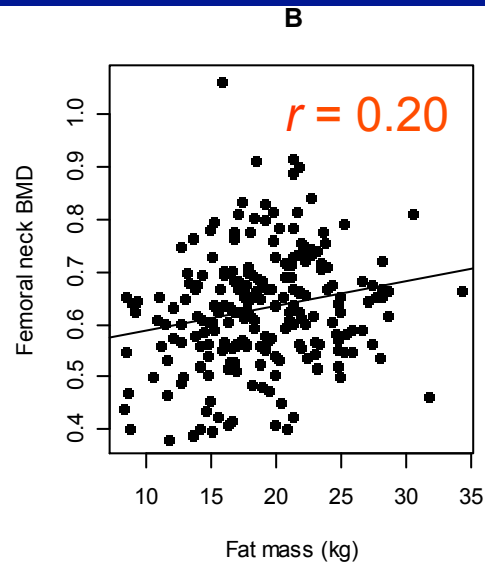
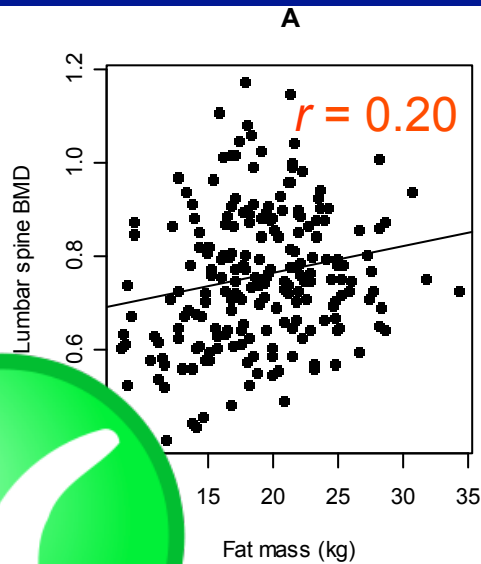
III. Results and discussion

1. Epidemiology (1/1985 - 12/2007)

1.1. Number of admitted patients



Tối đa hoá dữ liệu



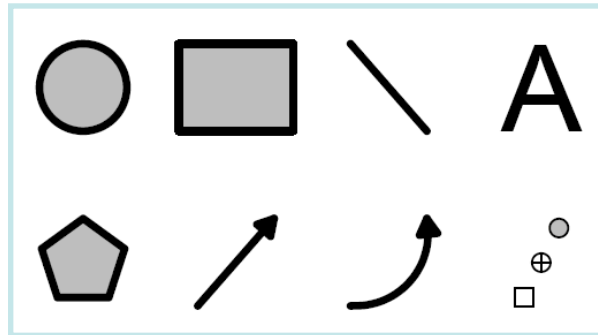
Một số biểu đồ chính trong khoa học

- Biểu đồ phân bố (histogram)
- Biểu đồ hộp (box plot)
- Biểu đồ thanh (bar plot)
- Biểu đồ tương quan (scatter plot)

Văn phạm biểu đồ trong R

Mô hình hoạ sĩ

- R mô phỏng theo hoạ sĩ
- Thêm các thành phần như thêm mực và vẽ màu



- Biểu đồ chỉ là những thành phần trên được xếp đặt một cách có ý nghĩa

Các package trong R

- R có nhiều hệ thống biểu đồ

graphics – truyền thống

`grid` – chức năng thấp

`lattice` – chức năng cao

ggplot2 – chất lượng cao

Các loại biểu đồ

<u>Dạng</u>	<u>R function ()</u>
Bar Chart	<code>barplot()</code>
Box & Whisker Plot	<code>boxplot()</code>
Dot Plot	<code>dotchartt()</code>
Histogram	<code>hist()</code>
Scatter Plot	<code>plot()</code>
Strip Chart	<code>stripchart()</code>

Trang trí

grid (nx, ny)	Thêm grid lines
axis (side n,)	Thêm axis
box (which=,)	Thêm box chung quanh biểu đồ
legend	Thêm legend
arrows (x, y) lines (x, y) points (x, y)	Thêm mũi tên, đường thẳng, điểm type = ("p", "b", "l", vv)
abline (a, b) abline (h= or v=)	Thêm đường biểu diễn (a = intercept, b = slope) h = horizontal, v = vertical
segments (x0, x1, y0, y1)	Thêm line segment(s) giữa các điểm
polygon (x, y)	Thêm polygon xác định bởi vector x và y
text (x, y, "note")	Thêm chữ trong biểu đồ

pch

- R có 25 plotting characters
- pch = định hình character
- Characters 19 – 20 can be filled with selected color
 - o 19 – Solid circle
 - o 20 – Bullet circle
- Characters 21: 25 can have selected fill and border colors
 - bg = Controls Border color
 - col= Controls fill

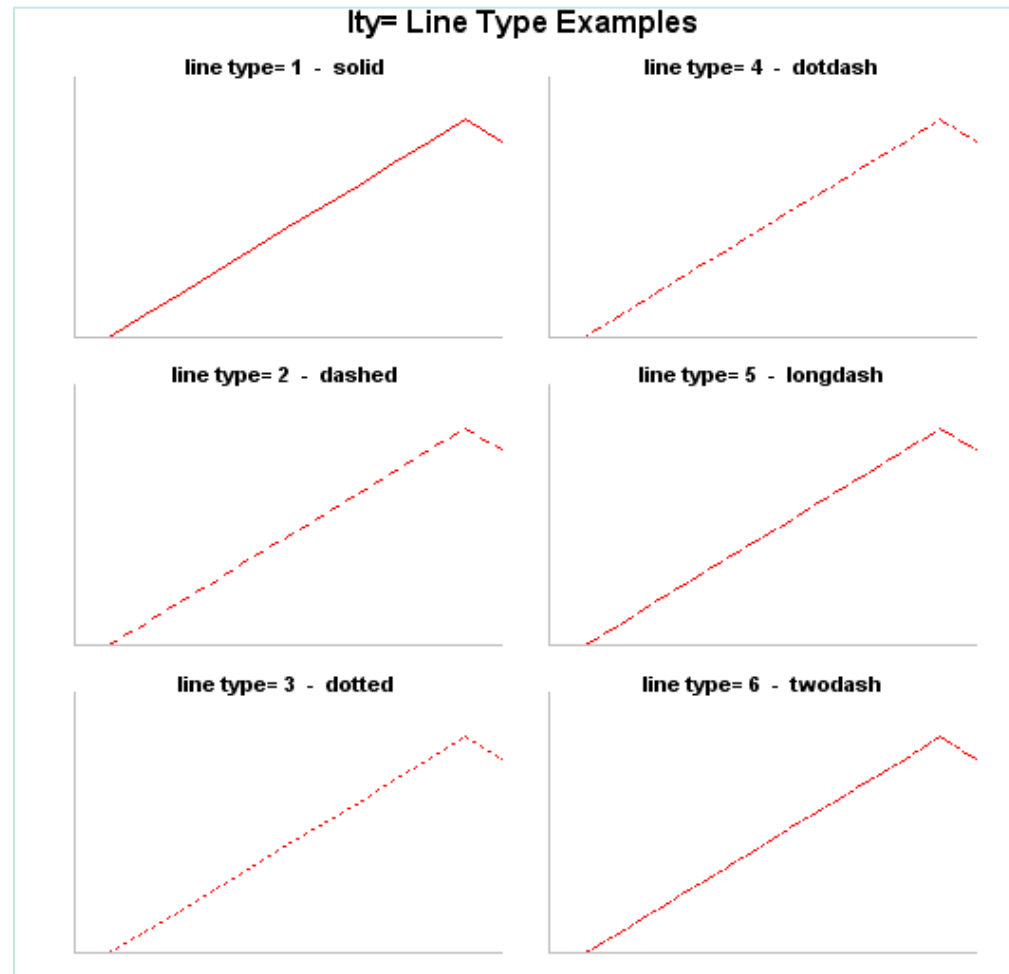
pch = Plot Characters

10 ⊕	20 ●	
9 ⊕	19 ●	
8 *	18 ◆	
7 ☒	17 ▲	
6 ▽	16 ●	
5 ◇	15 ■	25 ▼
4 ×	14 ☒	24 ▲
3 +	13 ⊗	23 ◆
2 △	12 ⊞	22 ■
1 ○	11 ⊗	21 ●

Symbols 19:25 can be filled with color.
19 and 20 are solid using col =
21:25 can be colored & filled with different colors
col= gives the border color
bg= gives the fill color

lty= Line Type Graphic Parameter

- R has 6 line types
- **lty** = specifies the line type
- **lty=** can be specified as integer or character string:
 - 0 – "blank"
 - 1 – "solid"
 - 2 – "dashed"
 - 3 – "dotted"
 - 4 – "dotdash"
 - 5 – "longdash"
 - 6 – "twodash"



**Làm quen với ngôn ngữ biểu đồ
trong R**

Các biến số

id	iid	sex	fx	timetofx	priorfx	hipfx	timetohipfx	fnbmd	lsbmd	age	weight	height
3	1	Male	0	0.55	0	0	0.55	1.08	1.46	73	98	175
8	2	Female	0	15.38	1	0	15.38	0.97	1.33	68	72	166
9	3	Male	0	5.06	0	0	5.06	1.01	1.49	68	87	184
10	4	Female	0	14.25	0	0	14.25	0.84	1.21	62	72	173
23	5	Male	0	15.07	0	0	15.07	0.81	1.14	61	72	173
24	6	Female	0	12.3	0	0	12.3	0.74	0.98	76	57	156

fnbmd: mật độ xương đo ở cổ xương đùi (g/cm^2)

lsbmd: mật độ xương đo ở cột sống thắt lưng (g/cm^2)

age: tuổi lúc tham gia nghiên cứu (năm)

weight: trọng lượng (kg) lúc tham gia nghiên cứu

height: chiều cao (cm) lúc tham gia nghiên cứu

Femoral neck BMD and age

sex

○ F

△ M

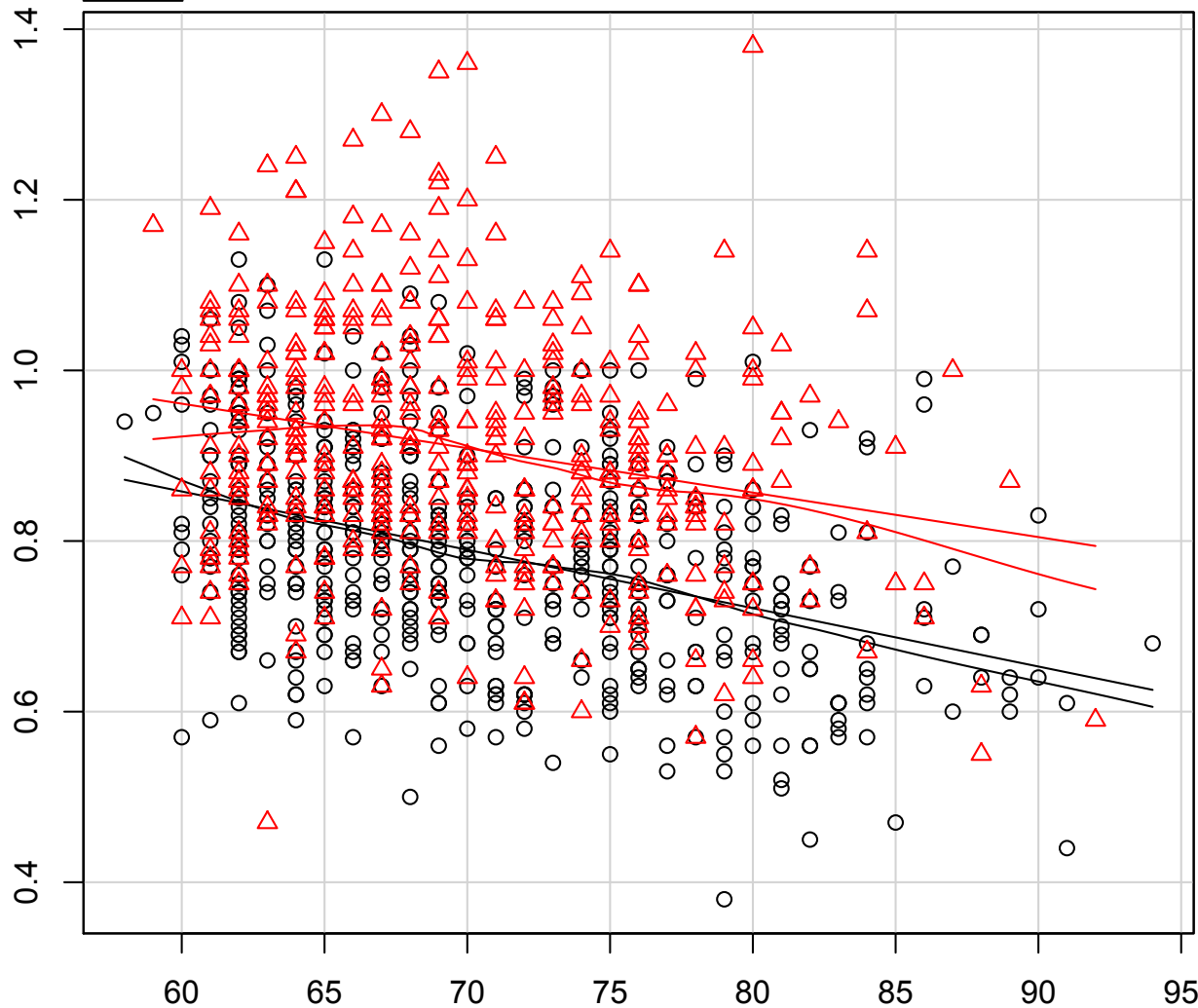
BMD

Age

ylim

main

xlim



Femoral neck BMD and age

sex

○ F

△ M

BMD

Age

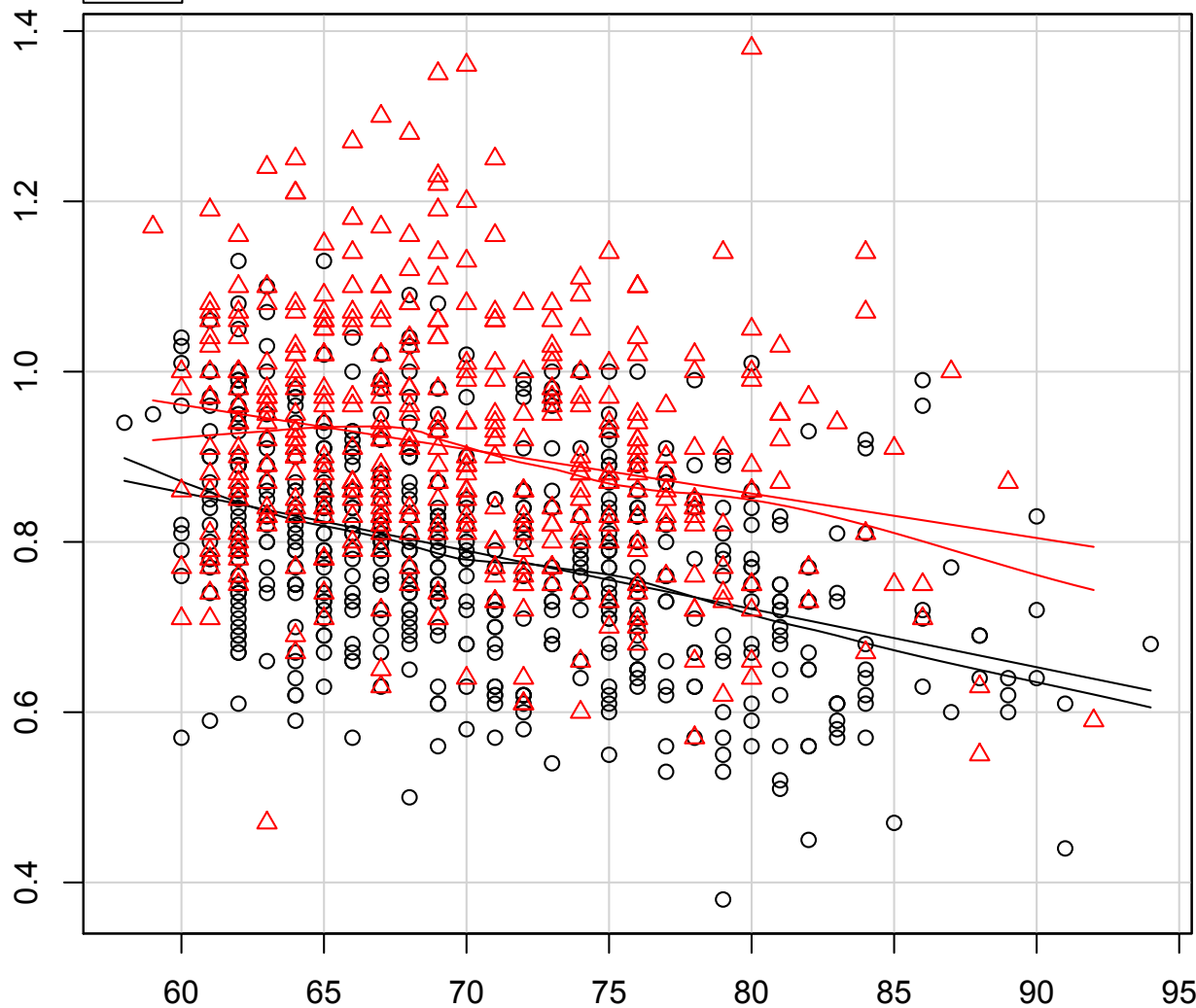
ylim

main

xlim

xlab

ylab



Femoral neck BMD and age

sex

○ F

△ M

BMD

Age

ylim

ylab

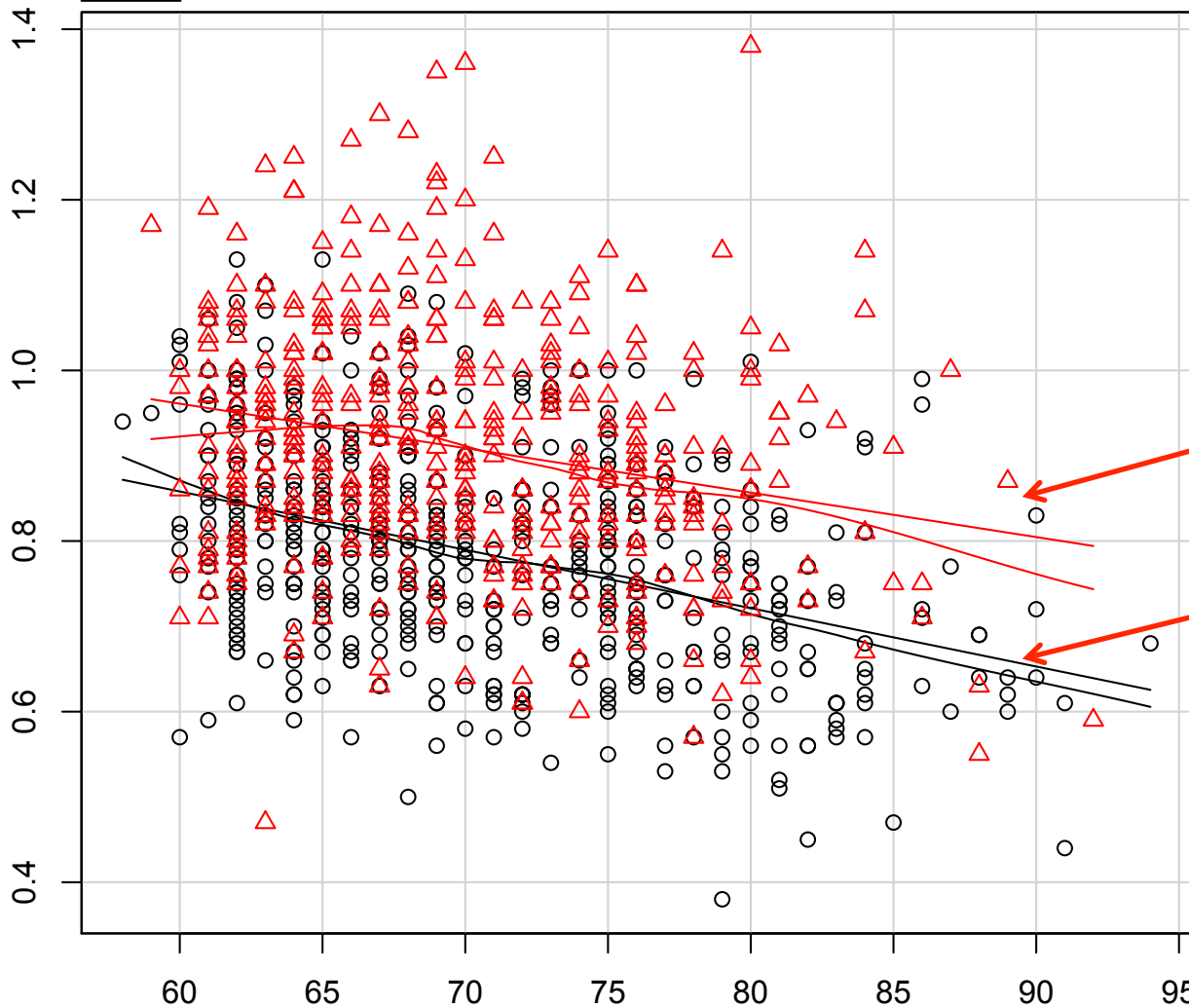
main

pch

abline

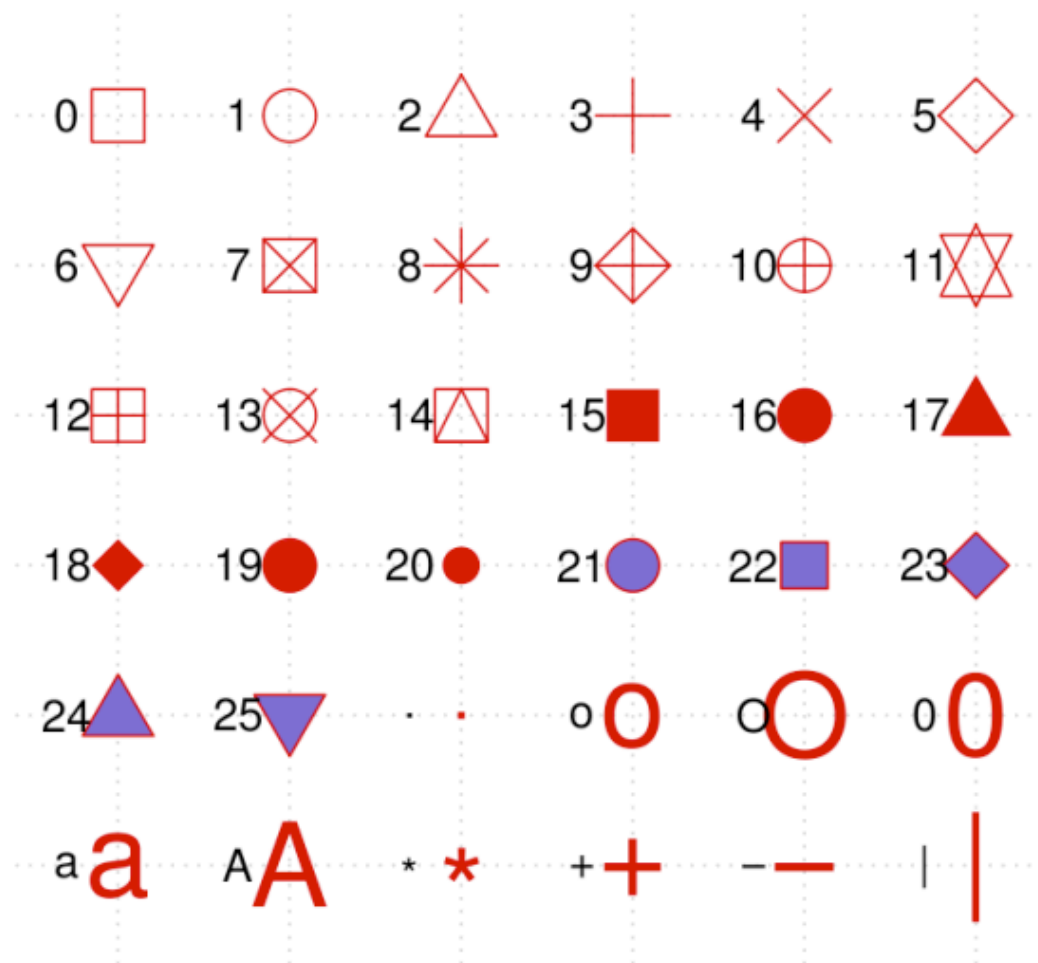
xlim

xlab



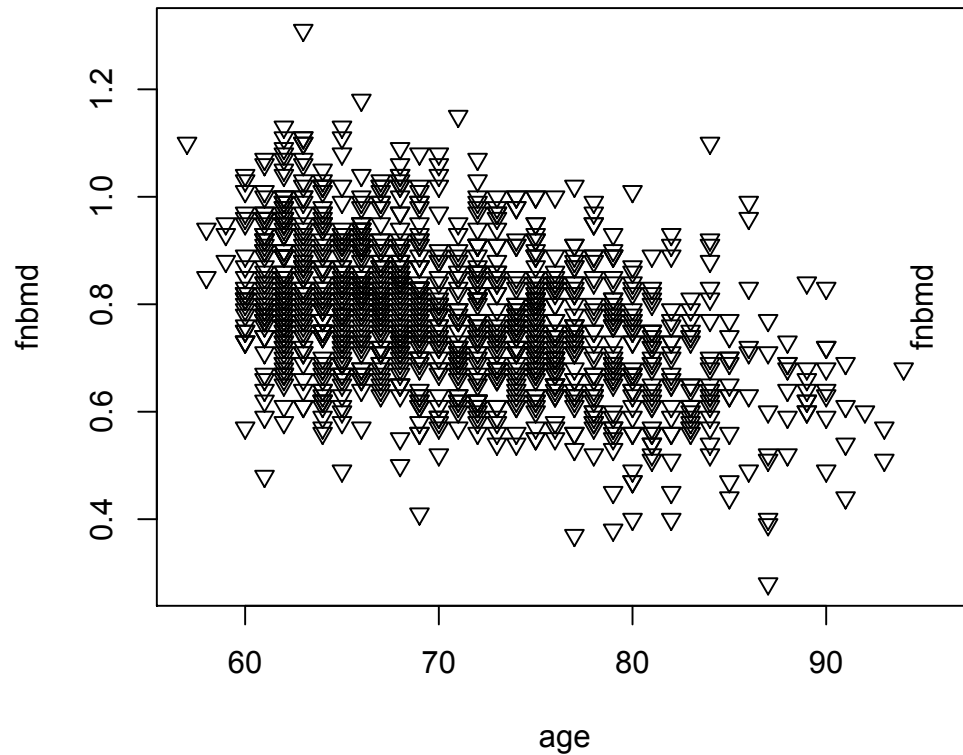
Kí hiệu điểm vẽ: *pch*

Plot symbols in R;
col = "red3 ", bg= " slateblue3 "

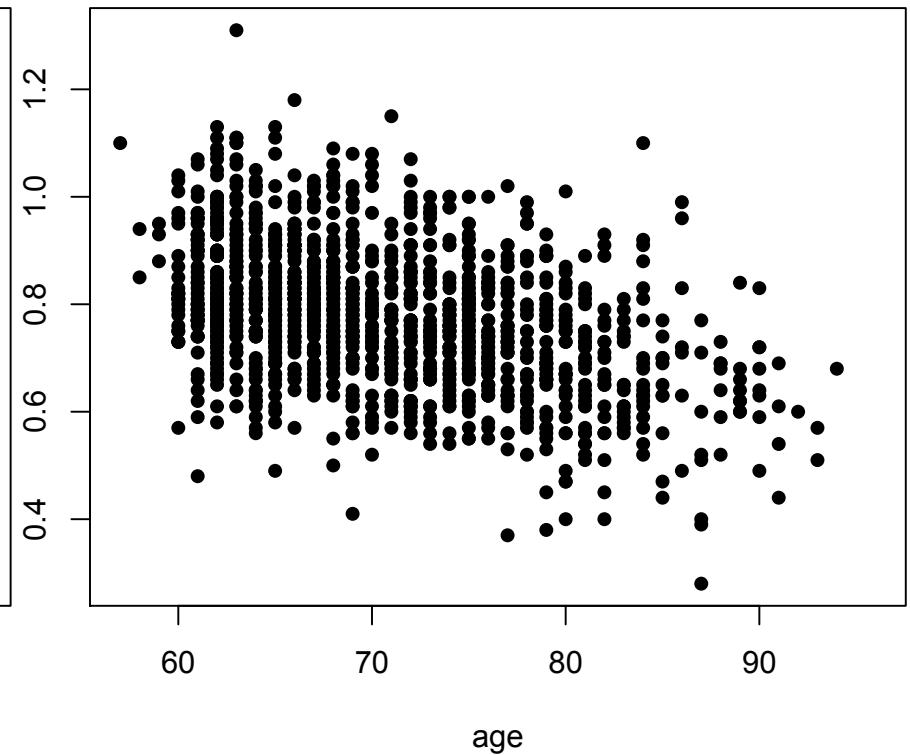


Tác động của pch

```
plot(fnbmd ~ age, pch=6)
```

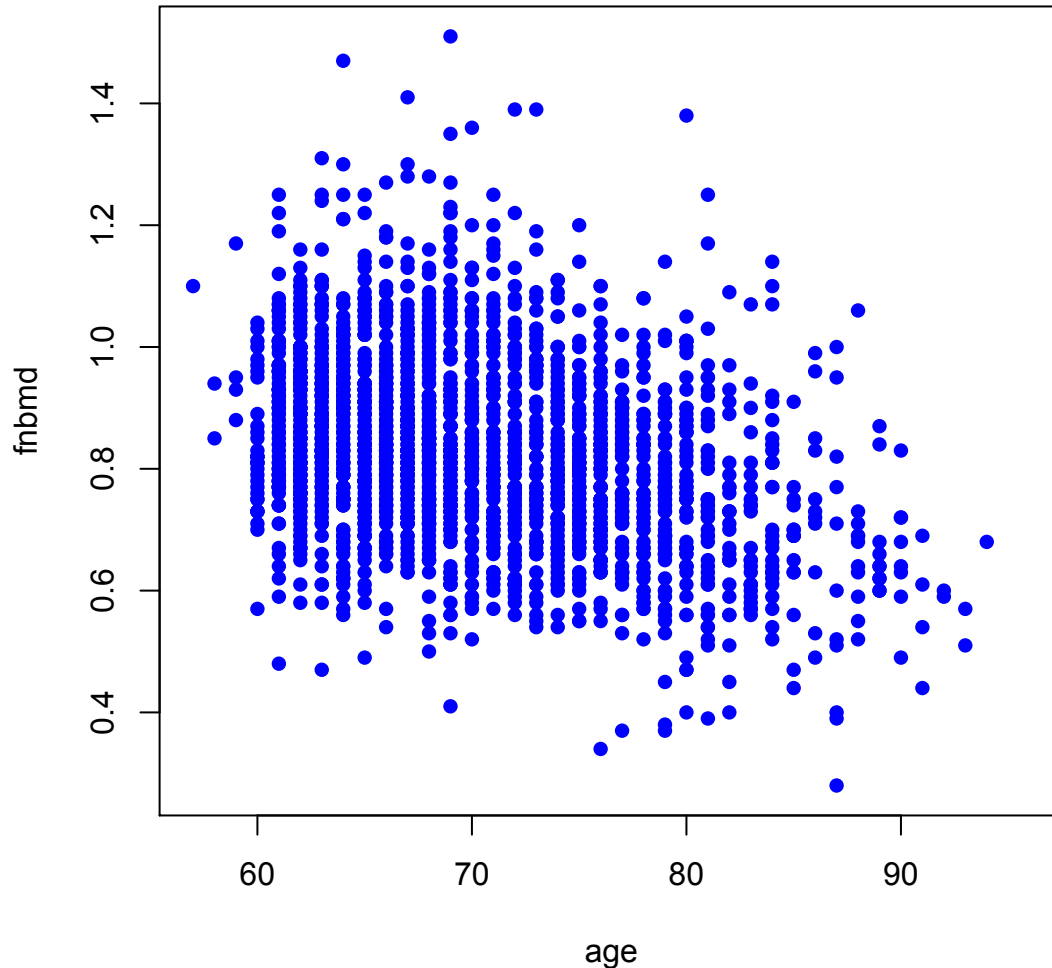


```
plot(fnbmd ~ age, pch=16)
```



Tác động của col

```
plot(fnbmd ~ age, pch=16, col="blue")
```

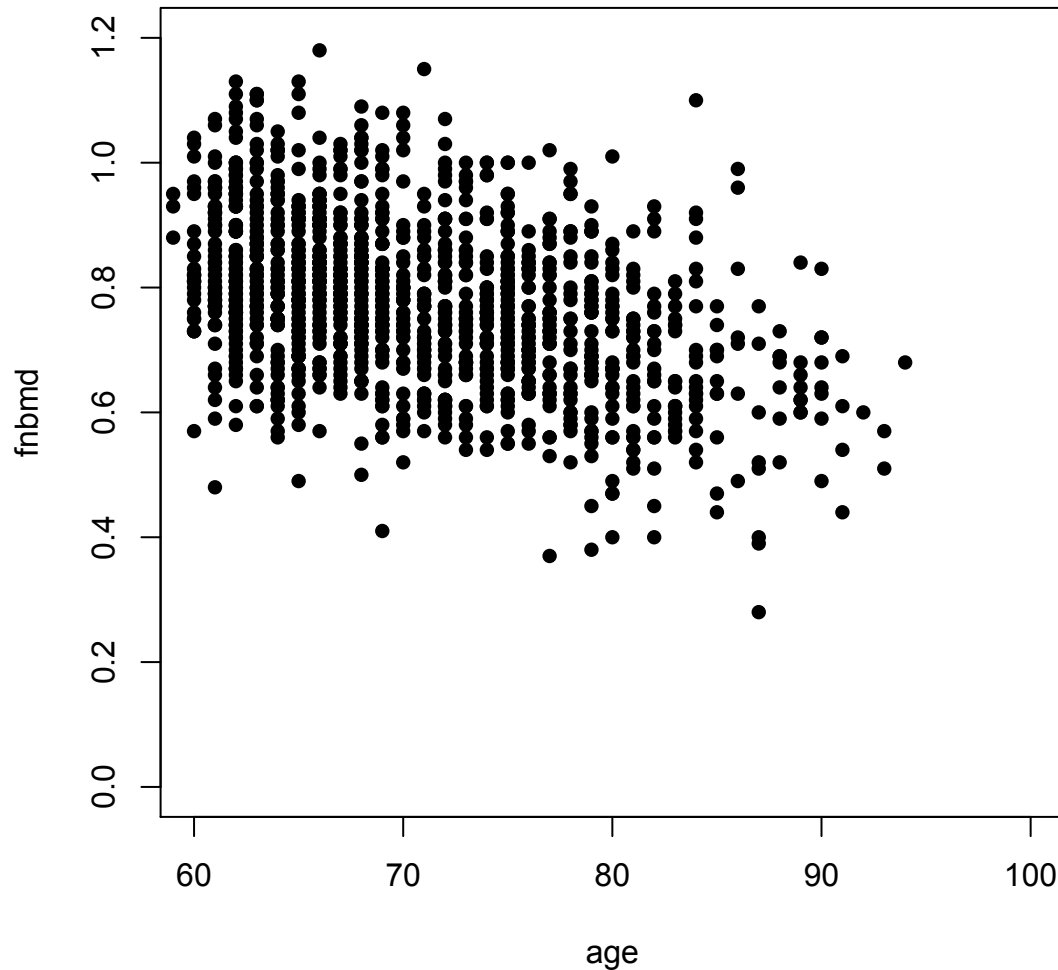


col=

- "black"
- "blue"
- "green"
- "red"
- "yellow"
- "pink"
- "orange"

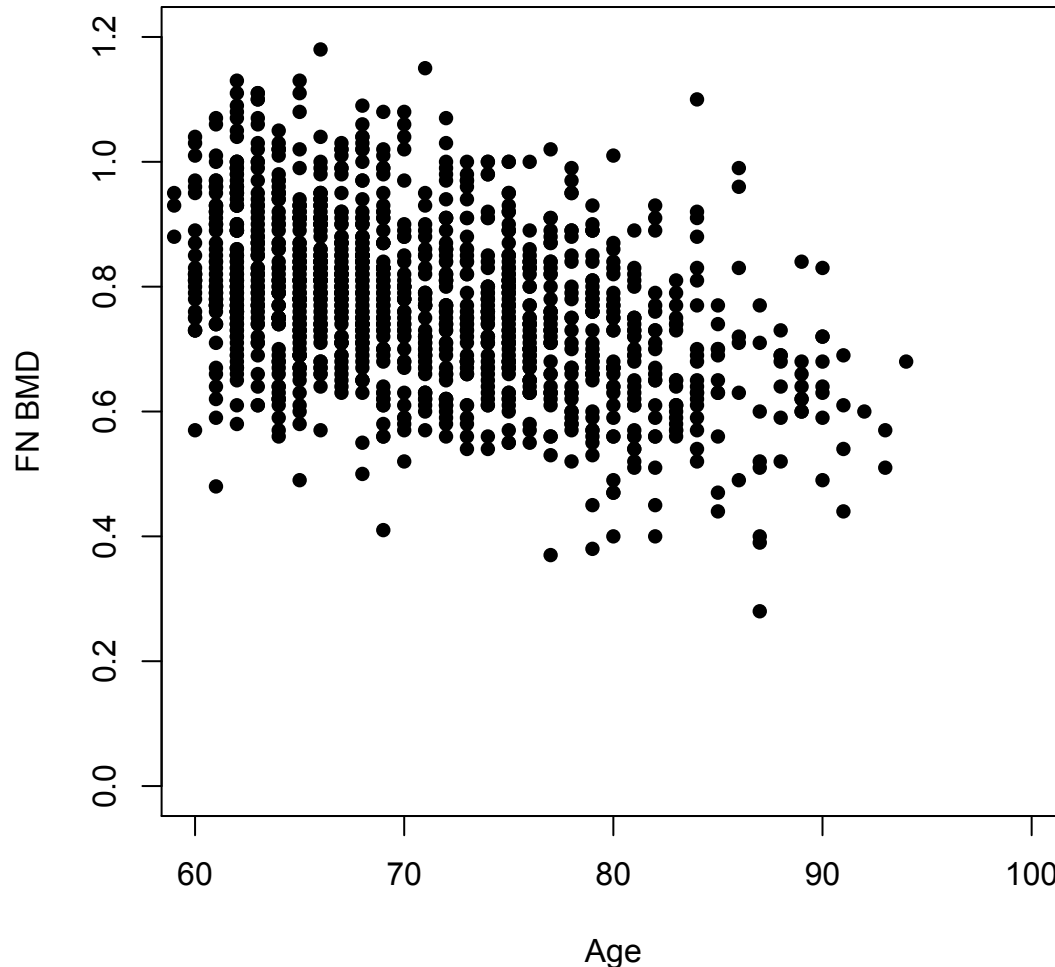
Tác động của xlim và ylim

```
plot(fnbmd ~ age, pch=16, ylim=c(0, 1.2),  
     xlim=c(60, 100))
```

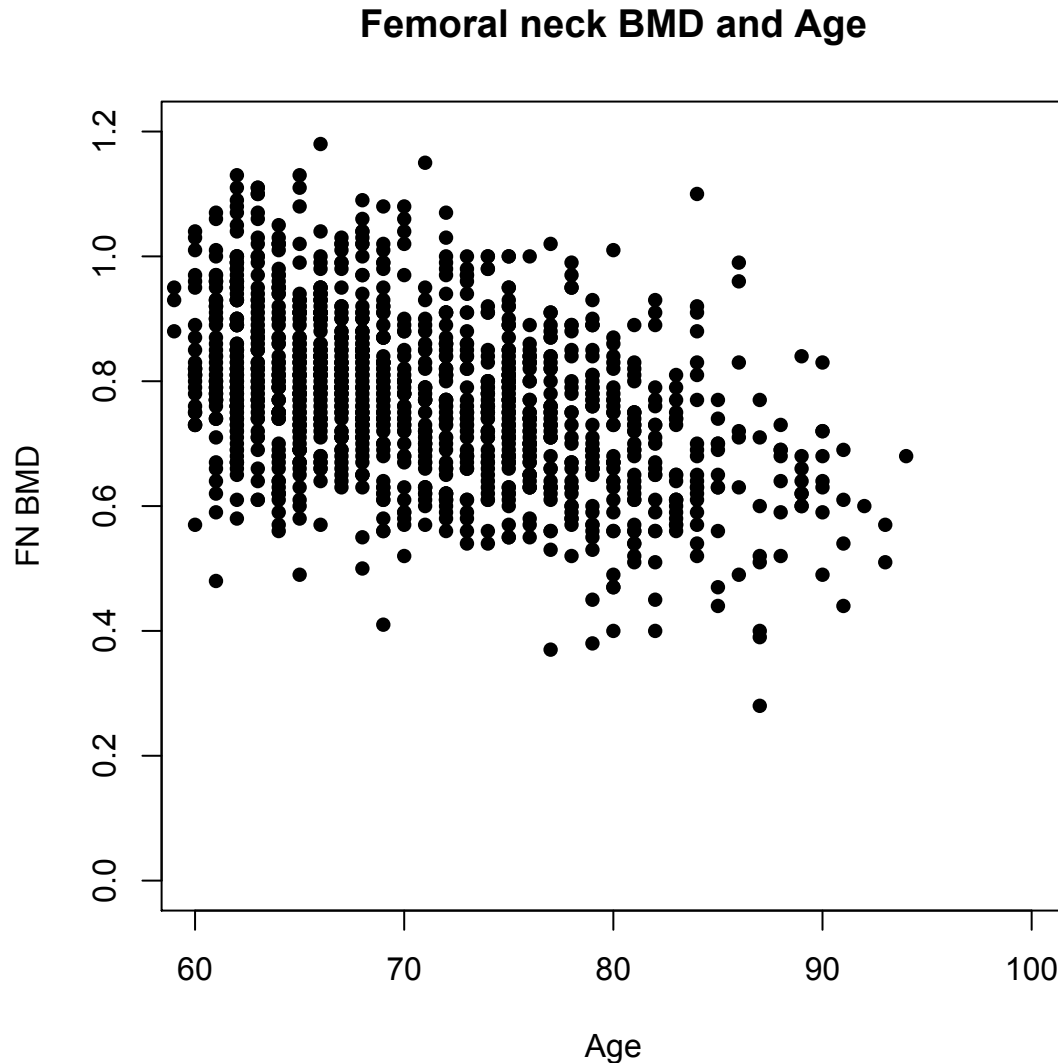


Tác động của xlab và ylab

```
plot(fnbmd ~ age, pch=16, ylim=c(0, 1.2),  
     xlim=c(60, 100), xlab="Age", ylab="FN BMD")
```



```
plot(fnbmd ~ age, pch=16, ylim=c(0, 1.2),  
     xlim=c(60,100), xlab="Age", ylab="FN BMD",  
     main="Femoral neck BMD and Age")
```



Thông số trong biểu đồ

Tựa đề

main

Tiêu đề trục hoành, trục tung: **xlab, ylab**

Khoảng giá trị trục hoành, tung: **xlim, ylim**

Màu

col

Kí hiệu (plotting character)

pch