

```
gender <- factor(c("M", "F", "M", "F", "F", "F", "F", "F", "F", "M", "M", "F", "M", "M", "F", "F"))
```

```
ethnicity <- factor (c("Asian", "Asian", "White", "White", "Hispanic", "White", "Hispanic",  
"Asian", "Asian", "White", "Hispanic", "Asian", "Asian", "Latin American", "White", "White"))
```

```
age <- c(29, 21, 27, 23, 21, 22, 24, 31, 24, 24, 24, 23, 24, 30, 23, 26)
```

```
lastname <- factor(c("H", "H", "R", "M", "L", "T", "F", "L", "J", "M", "L", "R", "A", "R", "S", "R"))
```

```
height <- c(72, 66, 69, 65, 62, 65, 59, 63, 64, 72, 69.5, 63, 65, 63, 68, 63)
```

```
datasummary <- data.frame(gender, ethnicity, age, lastname, height)
```

	gender	ethnicity	age	lastname	height
1	M	Asian	29	H	72.0
2	F	Asian	21	H	66.0
3	M	White	27	R	69.0
4	F	White	23	M	65.0
5	F	Hispanic	21	L	62.0
6	F	White	22	T	65.0
7	F	Hispanic	24	F	59.0
8	F	Asian	31	L	63.0
9	F	Asian	24	J	64.0
10	M	White	24	M	72.0
11	M	Hispanic	24	L	69.5
12	F	Asian	23	R	63.0
13	M	Asian	24	A	65.0
14	M	Latin American	30	R	63.0
15	F	White	23	S	68.0
16	F	White	26	R	63.0

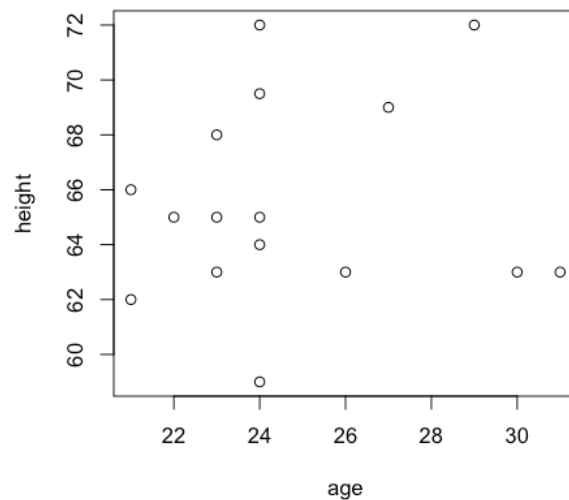
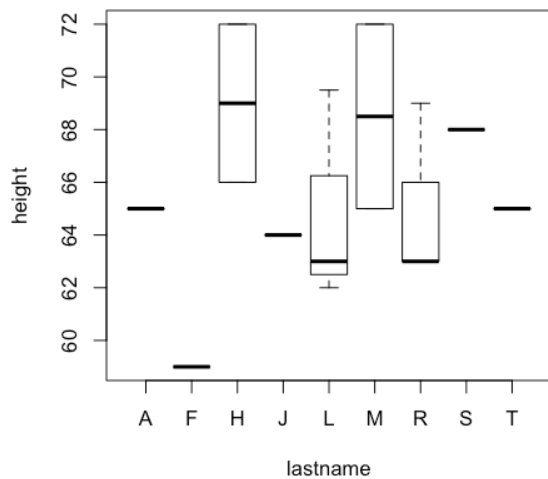
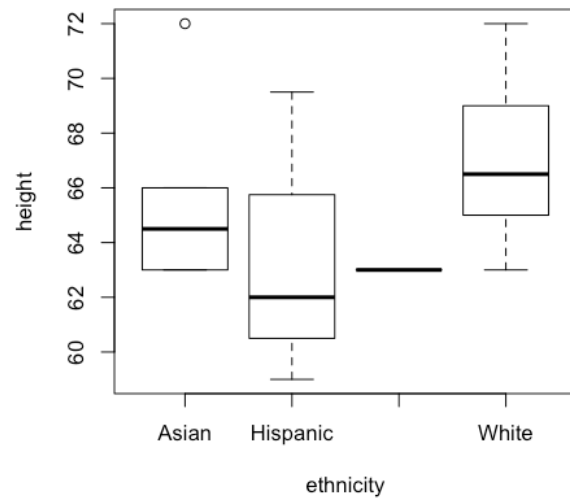
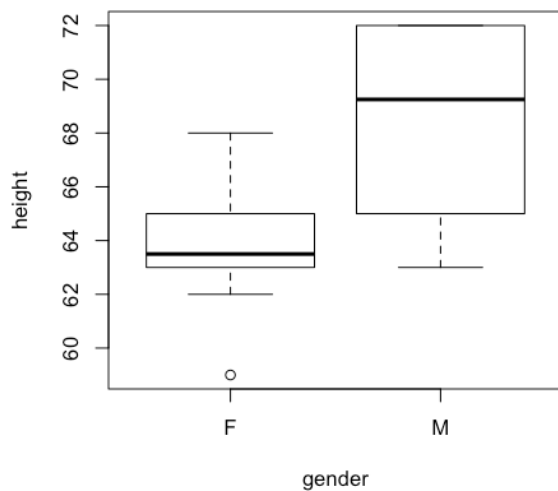
```
summary(datasummary)
```

gender	ethnicity	age	lastname	height
F:10	Asian :6	Min. :21.00	R :4	Min. :59.00
M: 6	Hispanic :3	1st Qu.:23.00	L :3	1st Qu.:63.00
	Latin American:1	Median :24.00	H :2	Median :65.00
	White :6	Mean :24.75	M :2	Mean :65.53
		3rd Qu.:26.25	A :1	3rd Qu.:68.25
		Max. :31.00	F :1	Max. :72.00
			(Other):3	

```

tiff(filename = "fig1", res = 100)
plot(height ~ gender, data = datasummary)
dev.off()
tiff(filename = "fig2", res = 100)
plot(height ~ ethnicity, data = datasummary)
dev.off()
tiff(filename = "fig3", res = 100)
plot(height ~ lastname, data = datasummary)
dev.off()
tiff(filename = "fig4", res = 100)
plot(height ~ age, data = datasummary)
dev.off()

```



Age didn't seem correlated with height according to the last figure.

There were not enough samples to make a conclusion about whether there's correlation between last name initials and height.

By inspection, the mean of male height was larger than the mean of female height. It's significant. As confirmed by the t-test below. It's consistent with the hypothesis and it's obvious.

As for ethnicity, the mean of white height was larger than others, but it's not significant. As confirmed by the t-test below. It might be consistent with the hypothesis but not obvious.

```
t.test(height[gender == "F"], height[gender == "M"])
```

Welch Two Sample t-test

```
data: height[gender == "F"] and height[gender == "M"]
t = -2.7255, df = 7.6701, p-value = 0.02707
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.5521539 -0.6811795
sample estimates:
mean of x mean of y
 63.80000  68.41667
```

```
t.test(height[ethnicity == "White"], height[ethnicity == "Asian"])
```

Welch Two Sample t-test

```
data: height[ethnicity == "White"] and height[ethnicity == "Asian"]
t = 0.77806, df = 9.9902, p-value = 0.4546
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.796127  5.796127
sample estimates:
mean of x mean of y
   67.0    65.5
```

```
t.test(height[ethnicity == "White"], height[ethnicity == "Hispanic"])
```

Welch Two Sample t-test

```
data: height[ethnicity == "White"] and height[ethnicity == "Hispanic"]
```

```
t = 1.0299, df = 2.7689, p-value = 0.3845
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-7.844744 14.844744
```

```
sample estimates:
```

```
mean of x mean of y
```

```
67.0      63.5
```

```
t.test(height[ethnicity == "White"], height[ethnicity == "Latin American"])
```

```
Error in t.test.default(height[ethnicity == "White"], height[ethnicity == "Latin American"] :  
not enough 'y' observations
```