Chicken soap

HUST Bioinformatics course series for undergraduates

Wei-Hua Chen (CC BY-NC 4.0)

06 April, 2021

Section 1 : outline

Outline

- how people in science see each other
- think in English
- think creative
- find your own truth
- 实力与包装: who i am underneath vs. whom people see in me

section 2 : contents

how people in science see each other

Figure 1: The most important illustration in science

lesions learned so far

- know your audiences (what they expect from you)
- be useful (in the long run)

think in English

怎么练习口语和语感？

1. Find a short article that you know every word
2. Read it aloud at a quiet place so that you can hear it
3. Do this everyday for **three years**

think creative

How?

1. Read a research paper
2. Find three limitations of the research
3. Address these issues with your own ideas
4. Do this once a week for **three years**

try find your own truth

未经调研，不要轻易相信别人结论，特别是科普文章。

There is a reproducibility crisis in scientific community:

- 70% of researchers failed to reproduce others' experiments
- 50% failed to reproduce their own

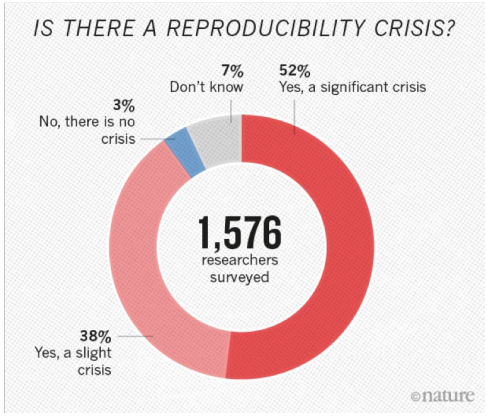
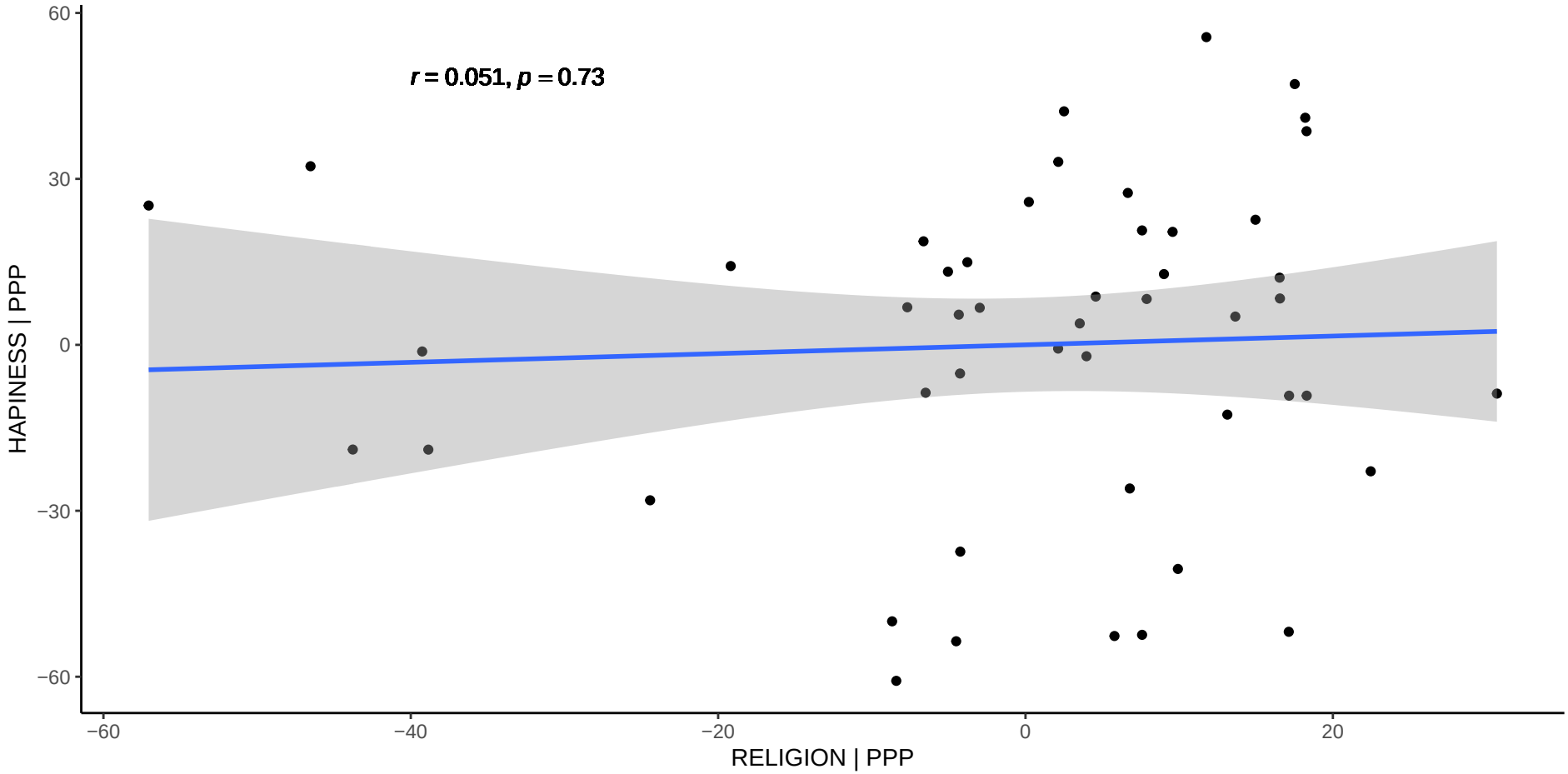


Figure 2: reproducibility crisis (Nature 2015 survey)

Religious and happiness

Are religious people happier? (Pew research 2019)

Religion, GDP and happiness.



The data

data/talk00/data.xlsx

```
head(rhp);
```

```
##                 happiness rel   ppp
## Switzerland        273.33  50 58551
## Austria            260.00  42 47250
## Iceland            260.00  57 46097
## Finland            256.67  53 41120
## Sweden             256.67  29 47922
## Canada             253.33  46 45553
```

Single factor analysis

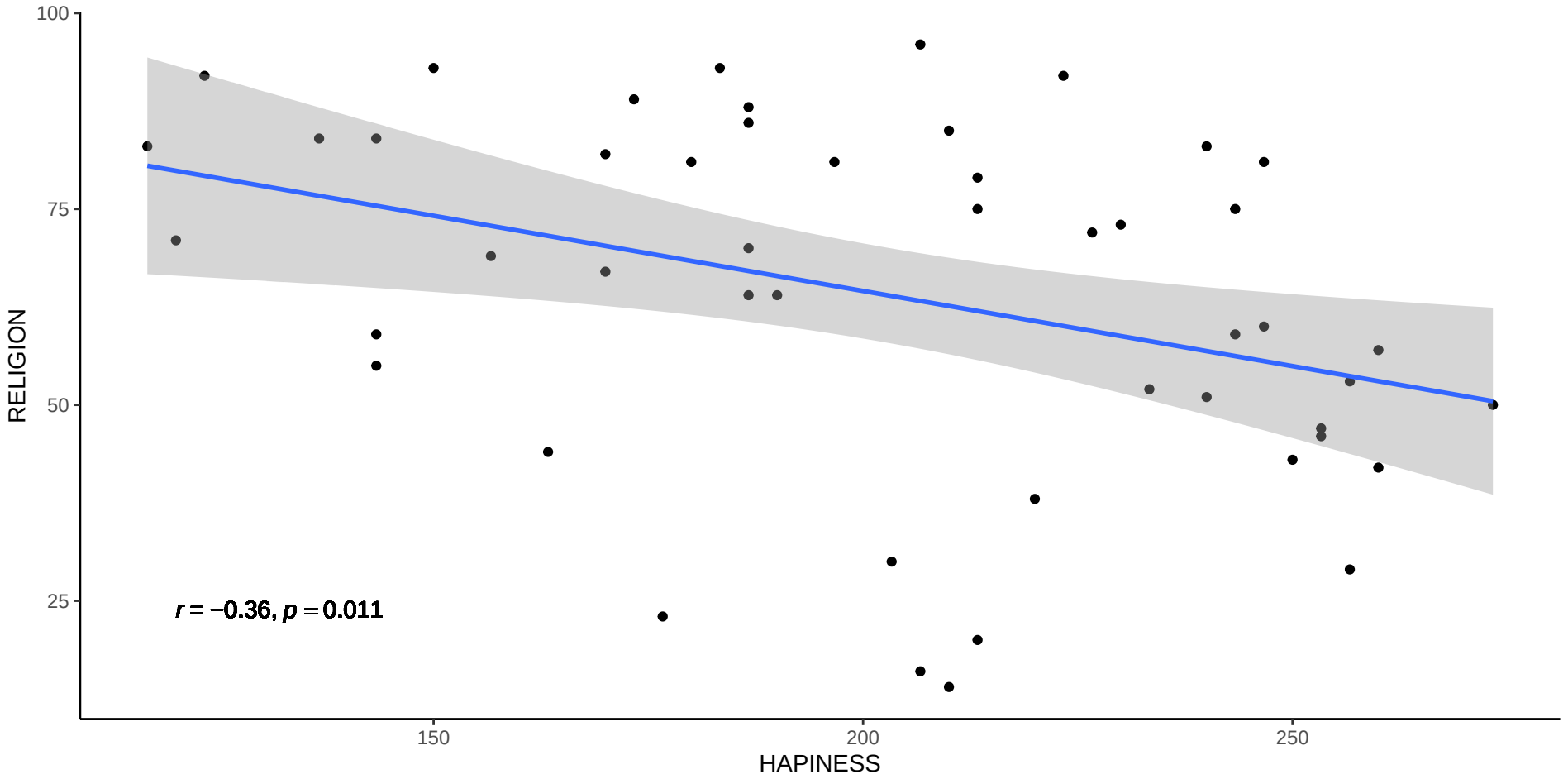
Happiness vs. Religion

```
c = cor.test( rhp$happiness, rhp$rel);
eq <- substitute(
  italic(r)~"="~r2*","~italic(p)==pvalue,
  list(
    r2 =  as.vector( format( c$estimate , digits = 2) ),
    pvalue =  as.vector( format( c$p.value , digits = 2) )
  ));
eq <- as.character(as.expression(eq));

m<-ggplot(rhp, aes(x=happiness, y=rel)) +
    geom_point() +
    labs(y="RELIGION", x = "HAPINESS")+
    theme_classic() +
    geom_smooth(method=lm) +
    geom_text( data = NULL,
               aes( x = 120, y = 25, label= eq, hjust = 0, vjust = 1),
               size = 4, parse = TRUE, inherit.aes=FALSE);
```

Plot

```
m
```



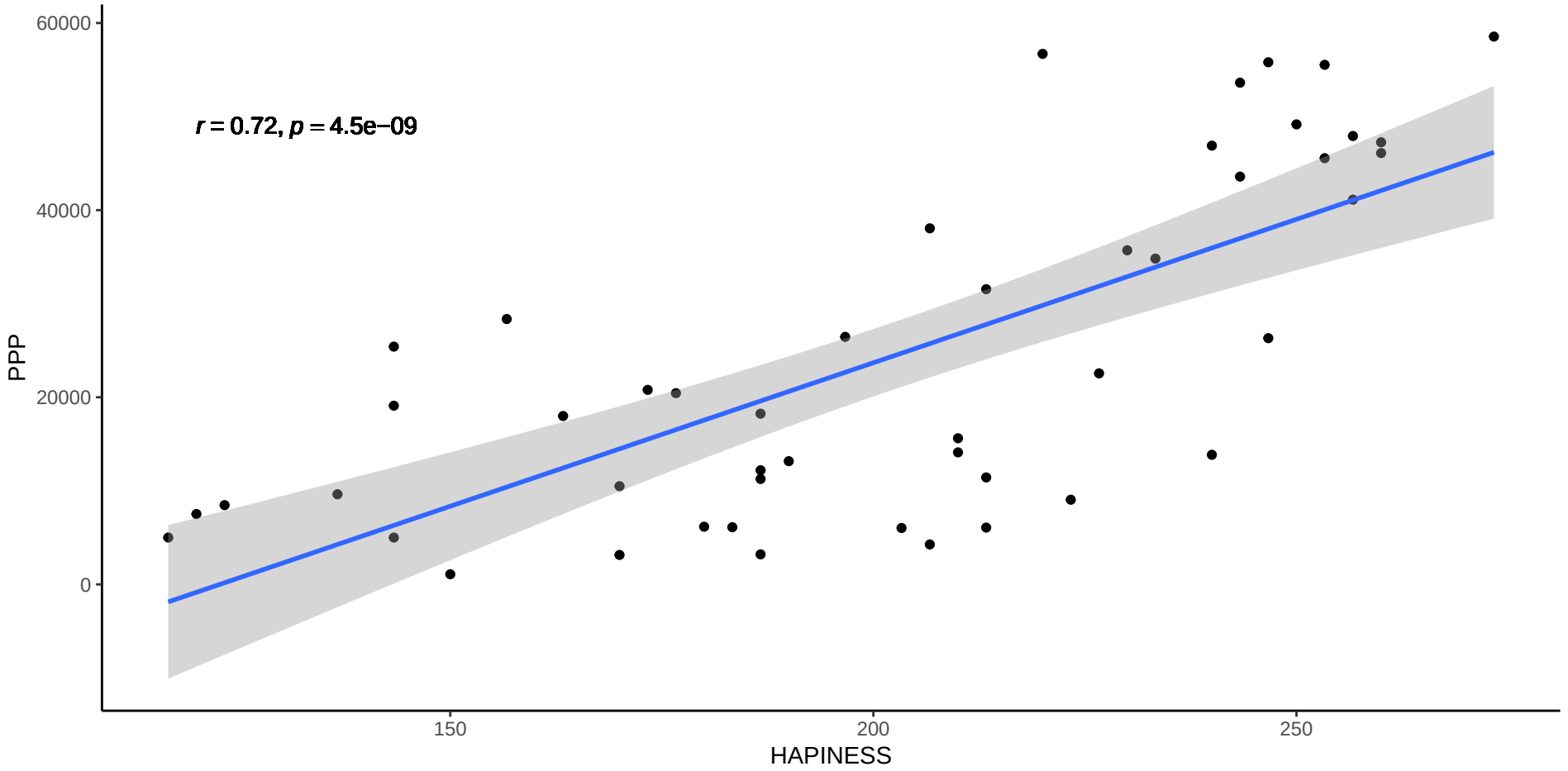
Happiness vs. ppp

```
c = cor.test( rhp$happiness, rhp$ppp);
eq <- substitute(
  italic(r)~"="~r2*","~italic(p)==pvalue,
  list(
    r2 =  as.vector( format( c$estimate , digits = 2) ),
    pvalue =  as.vector( format( c$p.value , digits = 2) )
  ));
eq <- as.character(as.expression(eq));

m2<-ggplot(rhp, aes(x=happiness, y=ppp)) +
    geom_point() +
    labs(y="PPP", x = "HAPINESS")+
    theme_classic() +
    geom_smooth(method=lm) +
    geom_text( data = NULL,
               aes( x = 120, y = 50000, label= eq, hjust = 0, vjust = 1),
               size = 4, parse = TRUE, inherit.aes=FALSE);
```

Plot

```
m2
```



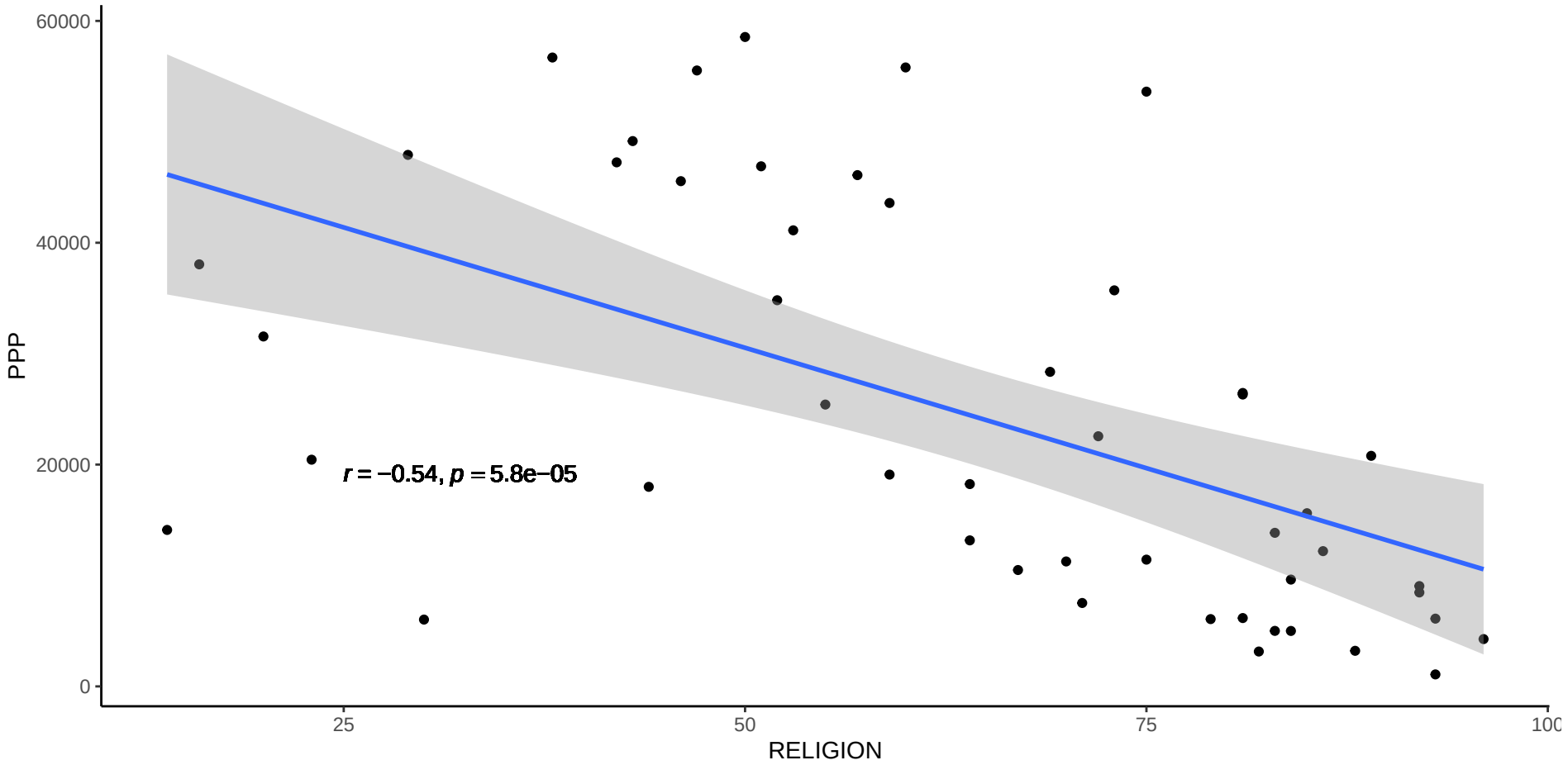
Religion vs. ppp

```
c = cor.test( rhp$rel, rhp$ppp);
eq <- substitute(
  italic(r)~"="~r2*","~italic(p)==pvalue,
  list(
    r2 =  as.vector( format( c$estimate , digits = 2) ),
    pvalue =  as.vector( format( c$p.value , digits = 2) )
  ));
eq <- as.character(as.expression(eq));

m3<-ggplot(rhp, aes(x=rel, y=ppp)) +
    geom_point() +
    labs(y="PPP", x = "RELIGION")+
    theme_classic() +
    geom_smooth(method=lm) +
    geom_text( data = NULL,
               aes( x = 25, y = 20000, label= eq, hjust = 0, vjust = 1),
               size = 4, parse = TRUE, inherit.aes=FALSE);
```

Plot

```
m3
```



Happiness vs. PPP vs. Religion

```
summary( glm( happiness ~ ppp + rel, data = rhp ) );
```

```
##
## Call:
## glm(formula = happiness ~ ppp + rel, data = rhp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -60.080  -15.469    5.794   19.668   54.697
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.544e+02  1.979e+01   7.801 5.83e-10 ***
## ppp         1.757e-03  2.853e-04   6.159 1.67e-07 ***
## rel         7.911e-02  2.286e-01   0.346    0.731
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 888.8269)
##
##     Null deviance: 85866  on 48  degrees of freedom
## Residual deviance: 40886  on 46  degrees of freedom
## AIC: 476.67
##
## Number of Fisher Scoring iterations: 2
```

信号传递理论：Michael Spence

观察：

- MBA 学生在进哈佛大学之前没什么了不起
- 出去就能比教授挣几倍、甚至十几倍的钱
- “这是为什么？”

结论：

- 教育具有信号传递的作用

Michael Spence, “Job Market Signaling”, Quarterly Journal of Economics, 1973

2001 Nobel price in Economics

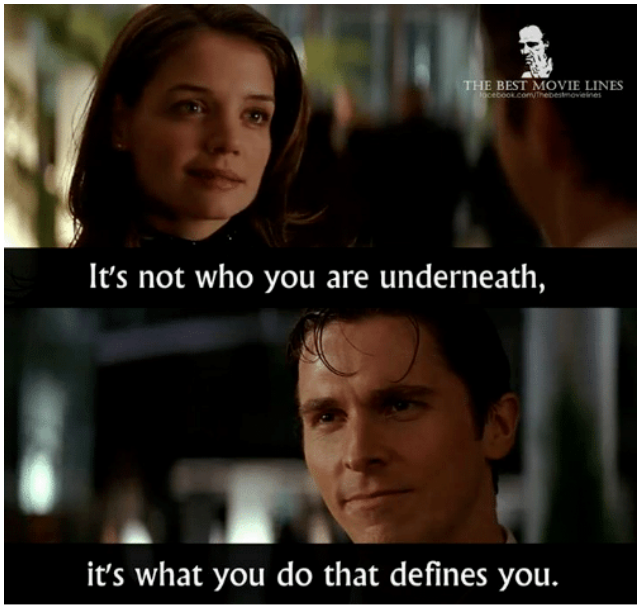
信号传递理论：实力与包装

信号传递理论的核心：动态不完全信息对策

1. 发送信息
2. 决策

Figure 3: Job interview cartoon

包装与实力



- Batman Begins 2005

Figure 4: The Batman Begins movie 2005

包装的素材

教育经历

- 名校
- 学科

科研经历

- 项目
- 成果
- 奖项

个人经历

- 干部
- 实习（领域知名企业）

Concluding remarks

- 有实力、会包装
- 不轻信
- 执着