Prepare the data and import it into R studio

Assign all variables to each vector respectively

准备好数据并且将它们导入R studio， 将每个变量抽取好

>Life\_Expectancy\_ <- read\_excel("Documents/10-Multiple Linear Regression-Life Expectancy-1400/Life Expectancy .xlsx")

> View(Life\_Expectancy\_)

> Country <- Life\_Expectancy\_$Country

> Year <- Life\_Expectancy\_$Year

> Status <- Life\_Expectancy\_$Status

> Life\_Expectancy <- Life\_Expectancy\_$`Life expectancy`

> Adult\_Mortality <- Life\_Expectancy\_$`Adult Mortality`

> Infant\_Deaths <- Life\_Expectancy\_$`infant deaths`

> Alcohol <- Life\_Expectancy\_$Alcohol

> Percentage\_Expenditure <- Life\_Expectancy\_$`percentage expenditure`

> Hepatitis\_B <- Life\_Expectancy\_$`Hepatitis B`

> Measles <- Life\_Expectancy\_$Measles

> BMI <- Life\_Expectancy\_$BMI

> Under\_Five\_Deaths <- Life\_Expectancy\_$`under-five deaths`

> Polio <- Life\_Expectancy\_$Polio

> Total\_Expenditure <- Life\_Expectancy\_$`Total expenditure`

> Diphtheria <- Life\_Expectancy\_$Diphtheria

> HIV\_AIDS <- Life\_Expectancy\_$`HIV/AIDS`

> GDP <- Life\_Expectancy\_$GDP

> Population <- Life\_Expectancy\_$Population

> Thinness\_1\_19 <- Life\_Expectancy\_$`thinness 1-19 years`

> Thinness\_5\_9 <- Life\_Expectancy\_$`thinness 5-9 years`

> Income\_Composition\_Of\_Resources <- Life\_Expectancy\_$`Income composition of resources`

> Schooling <- Life\_Expectancy\_$Schooling

1.分析的数据集中有哪些变量对人口预期寿命产生实际影响？

> princomp(na.omit(df1), cor = TRUE)

Call:

princomp(x = na.omit(df1), cor = TRUE)

Standard deviations:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10

2.47627994 1.70166494 1.35800847 1.26133551 1.10388771 0.93966186 0.89418263 0.81559306 0.72416677 0.71602901

Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17 Comp.18 Comp.19

0.64889113 0.62431525 0.58974445 0.57595140 0.45038140 0.35861318 0.26619046 0.19712976 0.04801655

19 variables and 1649 observations.

> fit <- princomp(na.omit(df1), cor = TRUE)

> summary(fit)

Importance of components:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8

Standard deviation 2.4762799 1.7016649 1.35800847 1.26133551 1.10388771 0.93966186 0.89418263 0.81559306

Proportion of Variance 0.3227349 0.1524033 0.09706247 0.08373512 0.06413516 0.04647181 0.04208224 0.03501011

Cumulative Proportion 0.3227349 0.4751382 0.57220068 0.65593580 0.72007096 0.76654277 0.80862501 0.84363512

Comp.9 Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16

Standard deviation 0.72416677 0.71602901 0.64889113 0.62431525 0.58974445 0.57595140 0.45038140 0.358613176

Proportion of Variance 0.02760092 0.02698408 0.02216104 0.02051419 0.01830519 0.01745895 0.01067597 0.006768601

Cumulative Proportion 0.87123604 0.89822012 0.92038116 0.94089535 0.95920053 0.97665948 0.98733545 0.994104048

Comp.17 Comp.18 Comp.19

Standard deviation 0.266190457 0.197129763 0.0480165526

Proportion of Variance 0.003729335 0.002045271 0.0001213468

Cumulative Proportion 0.997833382 0.999878653 1.0000000000

> loadings(fit)

Loadings:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11 Comp.12

Life\_Expectancy 0.325 0.195 0.240 0.145 0.126

Adult\_Mortality -0.217 -0.198 -0.385 -0.250 -0.150 -0.114 -0.164 -0.235 0.687

Infant\_Deaths -0.201 0.473 -0.142 -0.102 -0.144 -0.195

Alcohol 0.245 0.130 -0.109 -0.255 -0.236 -0.103 0.264 0.294 -0.299 -0.413 -0.360 -0.228

Percentage\_Expenditure 0.219 0.176 -0.289 -0.366 0.377 -0.207 -0.147

Hepatitis\_B 0.140 0.516 -0.152 -0.145 -0.204 0.414 -0.503 -0.147

Measles -0.111 0.322 -0.231 -0.248 -0.548 0.509 0.279 0.218

BMI 0.270 0.162 -0.261 0.136 -0.338 0.549 0.294 -0.468

Under\_Five\_Deaths -0.206 0.464 -0.145 -0.106 -0.168 -0.213

Polio 0.180 0.469 -0.188 -0.376 0.570 -0.320

Total\_Expenditure 0.129 -0.203 -0.188 0.897 -0.273 0.118

Diphtheria 0.188 0.509 -0.202 -0.102 0.189

HIV\_AIDS -0.141 -0.159 -0.107 -0.503 -0.345 0.139 0.234 0.326 0.369 -0.407

GDP 0.234 0.182 -0.273 -0.360 0.360 -0.182 -0.131

Population -0.111 0.418 -0.168 0.122 -0.494 -0.205 0.370 0.256

Thinness\_1\_19 -0.297 0.131 0.166 -0.141 0.335 0.166 0.288 0.182 0.206 -0.115

Thinness\_5\_9 -0.296 0.130 0.162 -0.150 0.327 0.152 0.318 0.203 0.202

Income\_Composition\_Of\_Resources 0.310 0.184 -0.101 0.310 0.214 0.118 0.282 0.179

Schooling 0.329 0.143 -0.121 0.320 0.183 0.158 0.158

Comp.13 Comp.14 Comp.15 Comp.16 Comp.17 Comp.18 Comp.19

Life\_Expectancy 0.169 0.836

Adult\_Mortality 0.178 0.163 0.209

Infant\_Deaths 0.283 0.203 -0.715

Alcohol -0.247 -0.317 -0.107

Percentage\_Expenditure -0.697

Hepatitis\_B -0.289 0.299

Measles -0.221 -0.146

BMI -0.273

Under\_Five\_Deaths 0.296 0.215 0.698

Polio -0.205 0.296

Total\_Expenditure

Diphtheria 0.535 -0.555

HIV\_AIDS 0.284

GDP 0.715

Population -0.463 -0.222

Thinness\_1\_19 -0.126 -0.707

Thinness\_5\_9 -0.127 0.705

Income\_Composition\_Of\_Resources 0.224 0.309 -0.651 -0.146

Schooling 0.157 0.703 -0.369

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11 Comp.12 Comp.13 Comp.14

SS loadings 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

Proportion Var 0.053 0.053 0.053 0.053 0.053 0.053 0.053 0.053 0.053 0.053 0.053 0.053 0.053 0.053

Cumulative Var 0.053 0.105 0.158 0.211 0.263 0.316 0.368 0.421 0.474 0.526 0.579 0.632 0.684 0.737

Comp.15 Comp.16 Comp.17 Comp.18 Comp.19

SS loadings 1.000 1.000 1.000 1.000 1.000

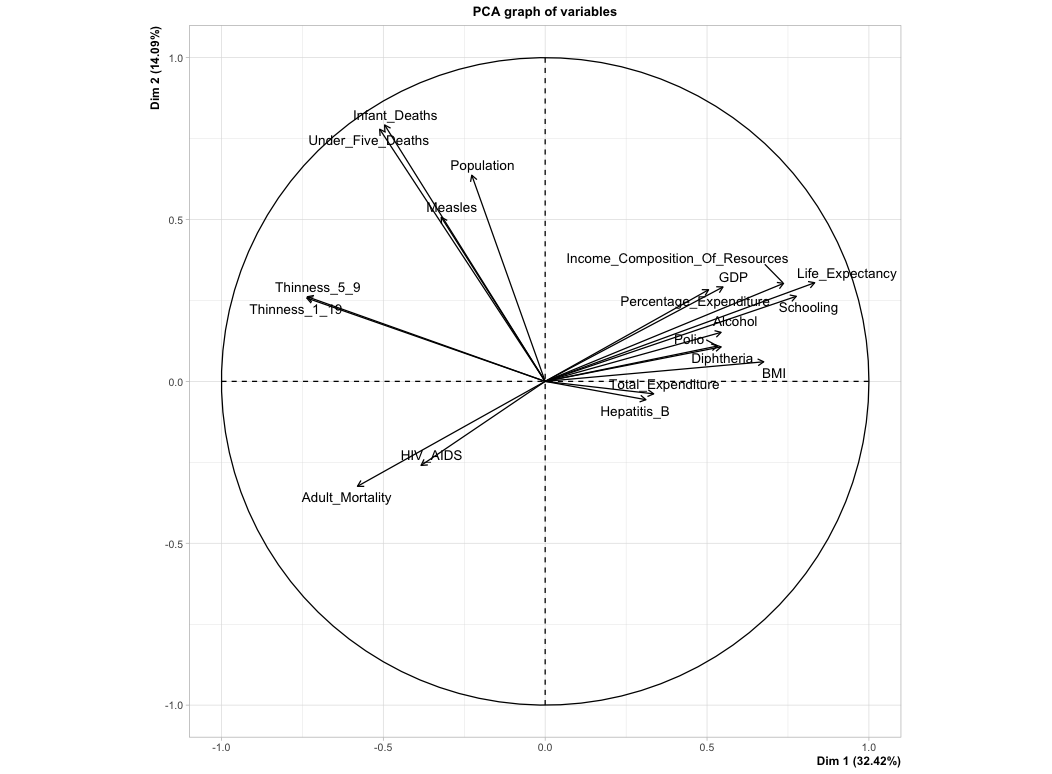
Proportion Var 0.053 0.053 0.053 0.053 0.053

Cumulative Var 0.789 0.842 0.895 0.947 1.000

> biplot(fit)

> library(FactoMineR)

> result <- PCA(df1)



我们可看出，

越靠近圆心的变量越产生实际影响，比如

HepatitisB-乙肝，

Total Expenditure – 医疗健康投入比列，

HIV， AIDS – 艾滋病

越靠近圆圈的变量产生实际影响越小，比如

Infant Deaths – 婴儿死亡数

Under 5 deaths – 小于5岁儿童的死亡数

2.人口预期寿命值低于65的国家是否应该增加其医疗保健支出以改善其平均寿命？

We first sort and filter the country that have less than 65 life expectancy In Excel and than copy past the data into a new file named “life expectancy less than 65”

我们首先将原表格里的人口预期寿命低于65 的国家和它们的数据筛选出来并且复制粘贴在另外一张Excel 表格上，并将它命名“life expectancy less than 65”

> library(readxl)

> Life\_Expectancy\_Less\_Than\_65 <- read\_excel("Documents/10-Multiple Linear Regression-Life Expectancy-1400/Life Expectancy Less Than 65.xlsx")

> View(Life\_Expectancy\_Less\_Than\_65)

> lifeexpectancy\_lower\_than\_65 <- Life\_Expectancy\_Less\_Than\_65$`Life expectancy`

> percentageexpenditure\_lower\_than\_65 <- Life\_Expectancy\_Less\_Than\_65$`percentage expenditure`

> model2 <- lm(lifeexpectancy\_lower\_than\_65 ~ percentageexpenditure\_lower\_than\_65)

> summary(model2)

Call:

lm(formula = lifeexpectancy\_lower\_than\_65 ~ percentageexpenditure\_lower\_than\_65)

Residuals:

Min 1Q Median 3Q Max

-20.4033 -3.9916 0.7099 5.0904 8.2114

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.669e+01 2.100e-01 269.997 <2e-16 \*\*\*

percentageexpenditure\_lower\_than\_65 4.036e-04 1.146e-03 0.352 0.725

---

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

Residual standard error: 5.621 on 849 degrees of freedom

Multiple R-squared: 0.0001461, Adjusted R-squared: -0.001032

F-statistic: 0.124 on 1 and 849 DF, p-value: 0.7248

> summary(model2)$coef

Estimate Std. Error t value Pr(>|t|)

(Intercept) 56.688616933 0.209960209 269.9969543 0.0000000

percentageexpenditure\_lower\_than\_65 0.000403595 0.001146025 0.3521696 0.7247985

根据结论我们可以看出, 在人口预期寿命低于65的国家里，医疗保健支出比例对人口寿命的影响微乎其微（0.000403595），我认为这些国家应该加强其他方面的建设与投入来提高总人口预期寿命而非将重点投入在其国民的医疗保健支出比例

3. 婴儿和成人死亡率如何影响人口预期寿命？

> lifeexpectancy <- Life\_Expectancy\_$`Life expectancy`

> df3 <- data.frame(lifeexpectancy, adultmortality, infantdeath)

> model3 <- lm(lifeexpectancy ~ adultmortality + infantdeath, data = df3)

> model3

Call:

lm(formula = lifeexpectancy ~ adultmortality + infantdeath, data = df3)

Coefficients:

(Intercept) adultmortality infantdeath

78.2260 -0.0525 -0.0115

> summary(model3)

Call:

lm(formula = lifeexpectancy ~ adultmortality + infantdeath, data = df3)

Residuals:

Min 1Q Median 3Q Max

-33.389 -2.529 1.315 3.943 17.025

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 78.2260313 0.2065916 378.65 <2e-16 \*\*\*

adultmortality -0.0524979 0.0009998 -52.51 <2e-16 \*\*\*

infantdeath -0.0114981 0.0010521 -10.93 <2e-16 \*\*\*

---

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

Residual standard error: 6.702 on 2925 degrees of freedom

(10 observations deleted due to missingness)

Multiple R-squared: 0.5051, Adjusted R-squared: 0.5048

F-statistic: 1493 on 2 and 2925 DF, p-value: < 2.2e-16

> summary(model3)$coef

Estimate Std. Error t value Pr(>|t|)

(Intercept) 78.22603130 0.2065916447 378.65051 0.000000e+00

adultmortality -0.05249792 0.0009997834 -52.50930 0.000000e+00

infantdeath -0.01149813 0.0010520741 -10.92901 2.800523e-27

根据结论我们可以看出婴儿和成人死亡率对预期人口寿命都有负相关影响，但成人死亡率（-0.05249792）较婴儿死亡率（-0.01149813）对人口预期寿命有更大的负影响

4.人口预期寿命与各变量间是正相关还是负相关

> model1 <- lm(Life\_Expectancy ~ Adult\_Mortality + Infant\_Deaths + Alcohol + Percentage\_Expenditure + Hepatitis\_B + Measles +BMI + Under\_Five\_Deaths + Polio + Total\_Expenditure + Diphtheria + HIV\_AIDS + GDP + Population + Thinness\_1\_19 + Thinness\_5\_9 + Income\_Composition\_Of\_Resources + Schooling)

> model1

> model1

Call:

lm(formula = Life\_Expectancy ~ Adult\_Mortality + Infant\_Deaths +

Alcohol + Percentage\_Expenditure + Hepatitis\_B + Measles +

BMI + Under\_Five\_Deaths + Polio + Total\_Expenditure + Diphtheria +

HIV\_AIDS + GDP + Population + Thinness\_1\_19 + Thinness\_5\_9 +

Income\_Composition\_Of\_Resources + Schooling)

> summary(model1)

Call:

lm(formula = Life\_Expectancy ~ Adult\_Mortality + Infant\_Deaths +

Alcohol + Percentage\_Expenditure + Hepatitis\_B + Measles +

BMI + Under\_Five\_Deaths + Polio + Total\_Expenditure + Diphtheria +

HIV\_AIDS + GDP + Population + Thinness\_1\_19 + Thinness\_5\_9 +

Income\_Composition\_Of\_Resources + Schooling)

Residuals:

Min 1Q Median 3Q Max

-17.0176 -2.0454 -0.0185 2.2260 11.9157

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.328e+01 7.358e-01 72.412 < 2e-16 \*\*\*

Adult\_Mortality -1.689e-02 9.473e-04 -17.828 < 2e-16 \*\*\*

Infant\_Deaths 9.369e-02 1.068e-02 8.776 < 2e-16 \*\*\*

Alcohol -5.435e-02 3.061e-02 -1.776 0.0760 .

Percentage\_Expenditure 3.777e-04 1.805e-04 2.093 0.0365 \*

Hepatitis\_B -5.582e-03 4.446e-03 -1.256 0.2095

Measles -8.617e-06 1.081e-05 -0.797 0.4253

BMI 3.350e-02 6.011e-03 5.573 2.92e-08 \*\*\*

Under\_Five\_Deaths -7.047e-02 7.728e-03 -9.119 < 2e-16 \*\*\*

Polio 7.836e-03 5.163e-03 1.518 0.1293

Total\_Expenditure 7.975e-02 4.074e-02 1.958 0.0505 .

Diphtheria 1.439e-02 5.938e-03 2.423 0.0155 \*

HIV\_AIDS -4.383e-01 1.788e-02 -24.519 < 2e-16 \*\*\*

GDP 1.383e-05 2.838e-05 0.487 0.6260

Population -6.917e-10 1.753e-09 -0.395 0.6931

Thinness\_1\_19 -8.670e-03 5.310e-02 -0.163 0.8703

Thinness\_5\_9 -5.123e-02 5.242e-02 -0.977 0.3286

Income\_Composition\_Of\_Resources 9.824e+00 8.340e-01 11.780 < 2e-16 \*\*\*

Schooling 8.783e-01 5.939e-02 14.789 < 2e-16 \*\*\*

---

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

Residual standard error: 3.596 on 1630 degrees of freedom

(1289 observations deleted due to missingness)

Multiple R-squared: 0.8347, Adjusted R-squared: 0.8329

F-statistic: 457.4 on 18 and 1630 DF, p-value: < 2.2e-16

> summary(model1)$coef

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.328156e+01 7.358126e-01 72.4118561 0.000000e+00

Adult\_Mortality -1.688853e-02 9.472913e-04 -17.8282337 4.305200e-65

Infant\_Deaths 9.369384e-02 1.067660e-02 8.7756282 4.200695e-18

Alcohol -5.434835e-02 3.060705e-02 -1.7756809 7.597207e-02

Percentage\_Expenditure 3.776826e-04 1.804691e-04 2.0927826 3.652285e-02

Hepatitis\_B -5.582500e-03 4.446185e-03 -1.2555707 2.094516e-01

Measles -8.617338e-06 1.080658e-05 -0.7974156 4.253258e-01

BMI 3.349806e-02 6.010676e-03 5.5730936 2.923353e-08

Under\_Five\_Deaths -7.047020e-02 7.728206e-03 -9.1185709 2.176104e-19

Polio 7.836379e-03 5.163246e-03 1.5177232 1.292781e-01

Total\_Expenditure 7.975222e-02 4.074157e-02 1.9575147 5.045741e-02

Diphtheria 1.438892e-02 5.938302e-03 2.4230697 1.549829e-02

HIV\_AIDS -4.382980e-01 1.787590e-02 -24.5189381 2.866052e-113

GDP 1.383274e-05 2.837729e-05 0.4874582 6.259992e-01

Population -6.916525e-10 1.752486e-09 -0.3946694 6.931385e-01

Thinness\_1\_19 -8.670229e-03 5.310244e-02 -0.1632736 8.703232e-01

Thinness\_5\_9 -5.122782e-02 5.241793e-02 -0.9772959 3.285677e-01

Income\_Composition\_Of\_Resources 9.824147e+00 8.339595e-01 11.7801248 8.430673e-31

Schooling 8.783350e-01 5.939038e-02 14.7891784 1.549772e-46

根据结论我们可以看出，

与人口预期寿命正相关的变量是带+号的，比如：

percentage expenditure医疗保健支出比列

BMI 身高体重指数

GDP 国民总产值

Schooling 受教育年龄

与人口预期寿命负相关的变量是带–号的，比如：

adult mortality, 承认死亡率

alcohol, 酗酒程度

hepatitis B, 乙肝

HIV\_AIDS, 艾滋