

810198540

گزارش پروژه اول آمار و
احتمال

معین کرمی

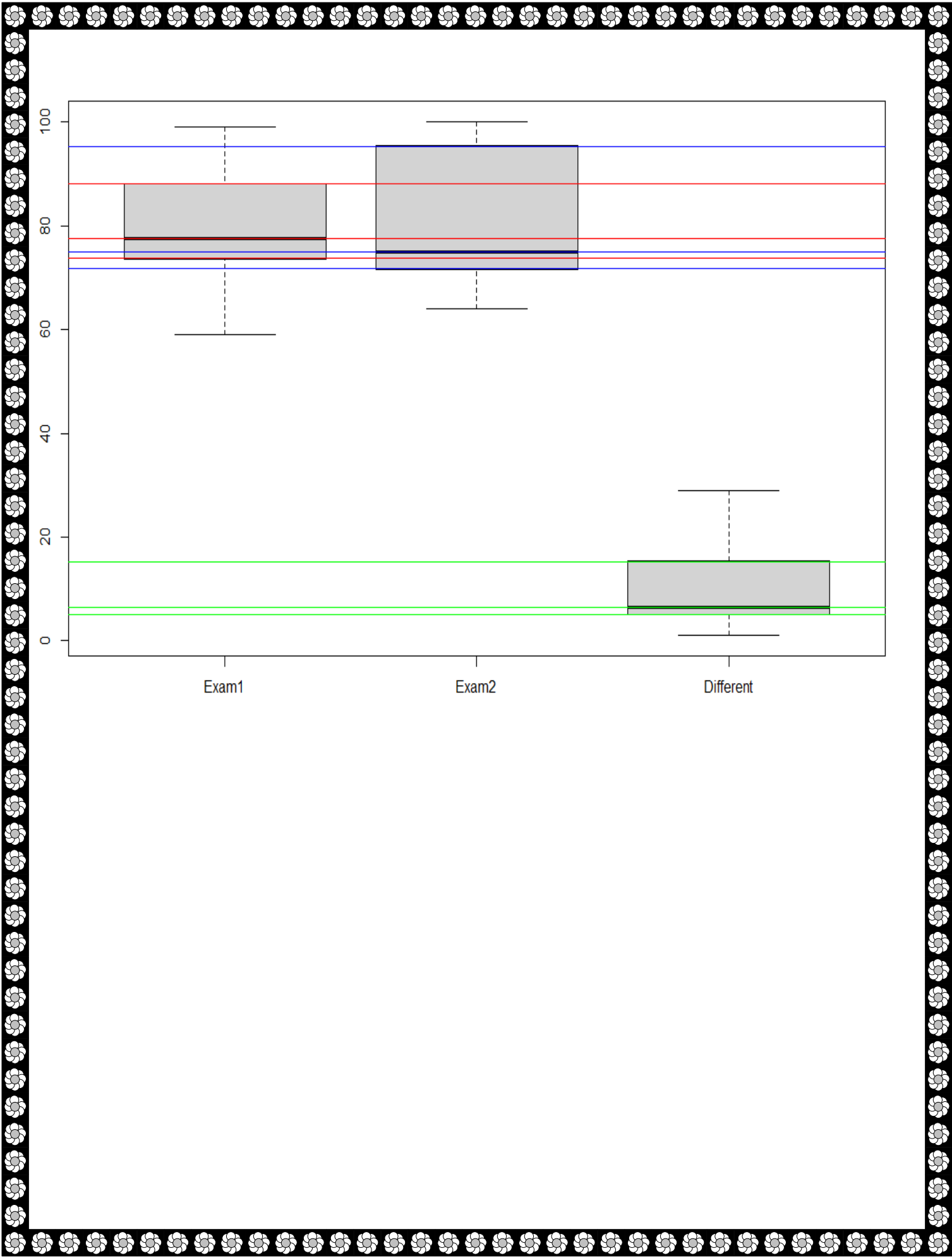
(بخش اول)

1- ابتدا با کمک تابع `read.csv()` اطلاعات را از فایل می‌خوانیم و در `grades` ذخیره می‌کنیم2- با کمک تابع `summary()` و `var()` اطلاعات خواسته شده را به دست می‌آوریم

Exam1	Exam2
Min. :59.00	Min. : 64.00
1st Qu.:73.75	1st Qu.: 71.75
Median :77.50	Median : 75.00
Mean :79.25	Mean : 80.55
3rd Qu.:88.00	3rd Qu.: 95.25
Max. :99.00	Max. :100.00
Vars. 102.30	Vars. 160.78

3- قدر مطلق تفاوت دو ستون را در `different` ذخیره می‌کنیم و با استفاده از تابع `c()` با `grades` ترکیب می‌کنیم و با کمک `boxplot()` نمودار مربوطه را کشیده و با کمک تابع `quantile()` مقدار دقیق چارک‌ها را به دست آورده و با کمک `abline()` این مقادیر را در نمودار مشخص می‌کنیم.

	1 st Qu	2 nd Qu	3 rd Qu
Exam1	73.75	77.5	88.00
Exam2	71.75	75.0	95.25
Different	5.00	6.5	15.25



بخش دوم)

۱- با کمک تابع `na.string()` مقدار NA را به جای مقادیر خالی قرار می دهیم.

۲- با کمک تابع `colnames()` می توان اسم ستون ها را چاپ کرد.

Column names : age, workclass, education, martial.status, occupation
relationship, race, sex, native.country, income

Head :

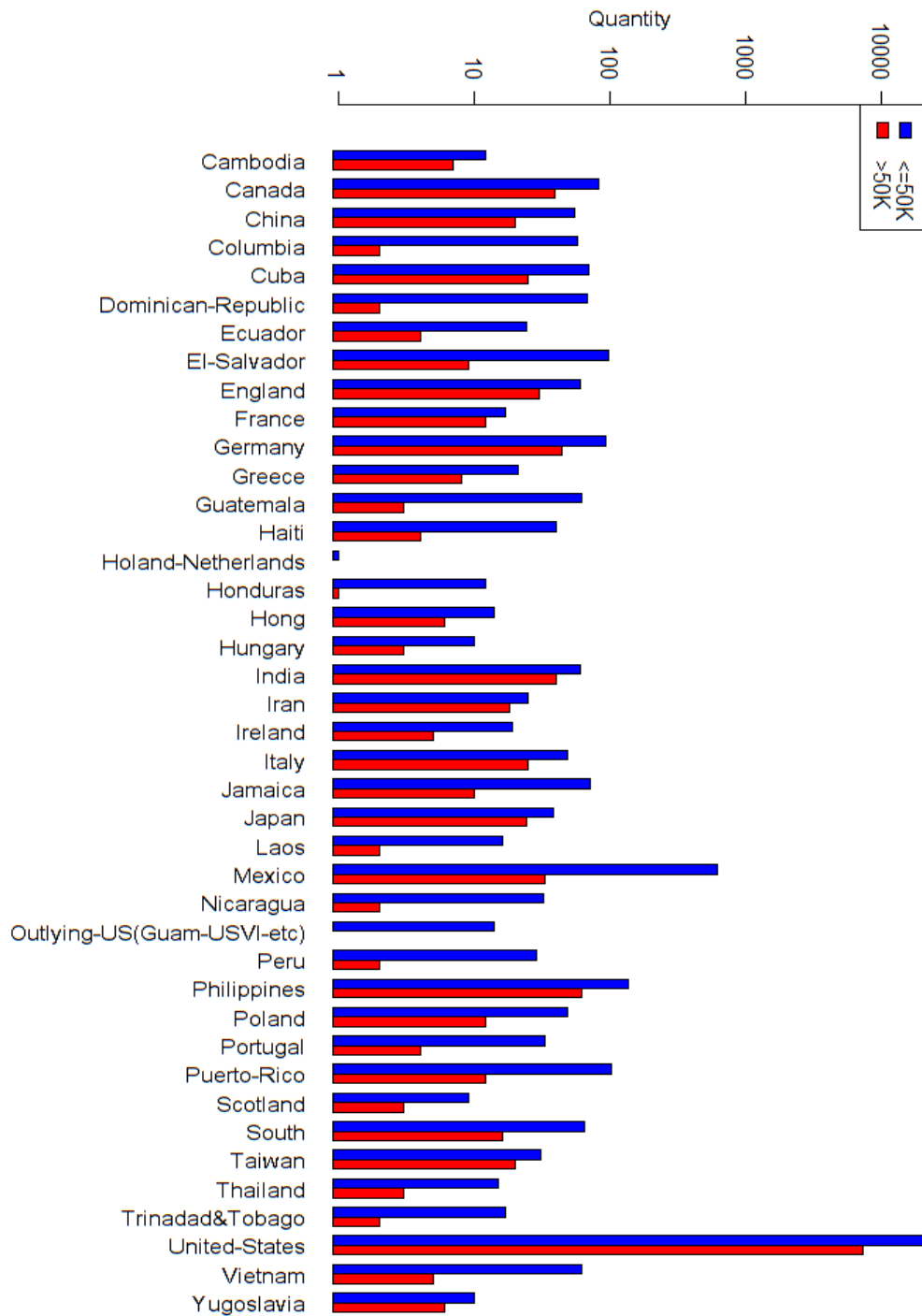
age	workclass	education	martial.status	occupation	relationship	race	sex	native.country	income
39	State-gov	Bachelors	Never-married	Adm-clerical	Not-in-family	White	Male	United-States	<=50K
50	Self-emp-not-inc	Bachelors	Married-civ-spouse	Exec-managerial	Husband	White	Male	United-States	<=50K
38	Private	HS-grad	Divorced	Handlers-cleaners	Not-in-family	White	Male	United-States	<=50K
53	Private	11th	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	United-States	<=50K
28	Private	Bachelors	Married-civ-spouse	Prof-specialty	Wife	Black	Female	Cuba	<=50K
37	Private	Masters	Married-civ-spouse	Exec-managerial	Wife	White	Female	United-States	<=50K

۳-

Relationship table:

Husband	Not-in-family	Other-relative	Own-child	Unmarried	Wife
13193	8305	981	5068	3446	1568

Income and native.country table grouped bar plot :



۴- با استفاده از \$ می‌توانیم به یک ستون خاص از دیتا فریم دسترسی داشته باشیم. بهتر است اسم دیتا فریم خود را data نگذاریم برای همین از یک دیتا فریم کمکی به اسم outcome استفاده می‌کنیم.

