МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

МЕЖГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Финансы и бухгалтерский учёт»

Отчёт по лабораторной работе № 5

СБОР И ПОДГОТОВКА ДАННЫХ

Выполнил

Студент гр. АСОИ-181

Остапенко А. К.

Проверила

Ливинская В. А.

Могилёв, 2021

*Цель работы:* исследование наборов данных, представленных на портале data.worldbank.org средствами R

Вариант 12

Тематика – Infrastructure (Инфраструктура)

Показатель – IS.SHP.GCNW.XQ (Liner shipping connectivity index (maximum value in 2004 = 100) – Индекс связности линейных перевозок (макс. значение в 2004 г. = 100%)). Показывает, насколько хорошо страны подключены к глобальным сетям судоходства на основании состояния их сектора морского транспорта.

**Выполнение**

**Получим данные из мирового банка данных с помощью пакета WDI**

*infr <- WDI(indicator = "IS.SHP.GCNW.XQ", country = "all", start = 2015, end = 2020, extra = TRUE)*

**Сделаем выборки по уровню дохода стран на 2015 год.**

*infr\_hight\_income <- subset(infr, income == "High income" & year == 2015)*

*infr\_low\_income <- subset(infr, income == "Low income" & year == 2015)*

*infr\_low\_mid\_income <- subset(infr, income == "Lower middle income" & year == 2015)*

*infr\_up\_mid\_income <- subset(infr, income == "Upper middle income" & year == 2015)*

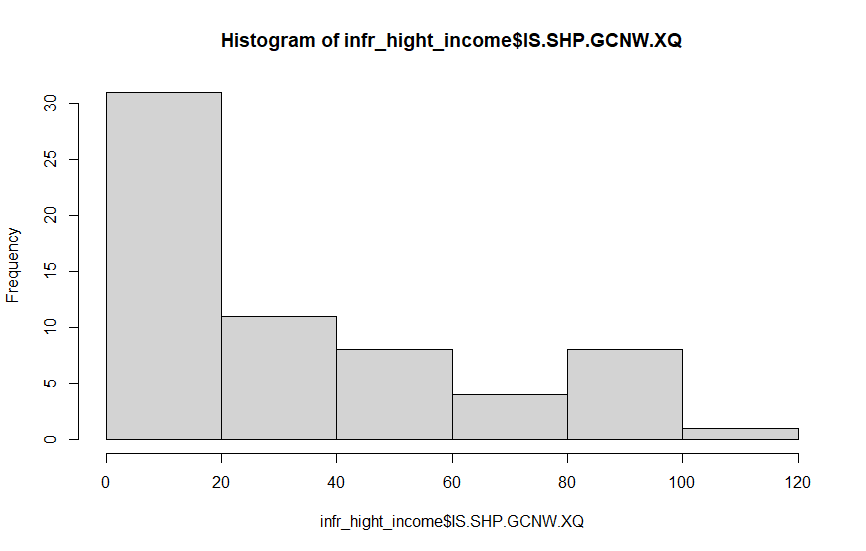
**Для каждой выборки построим гистограмму распределения**

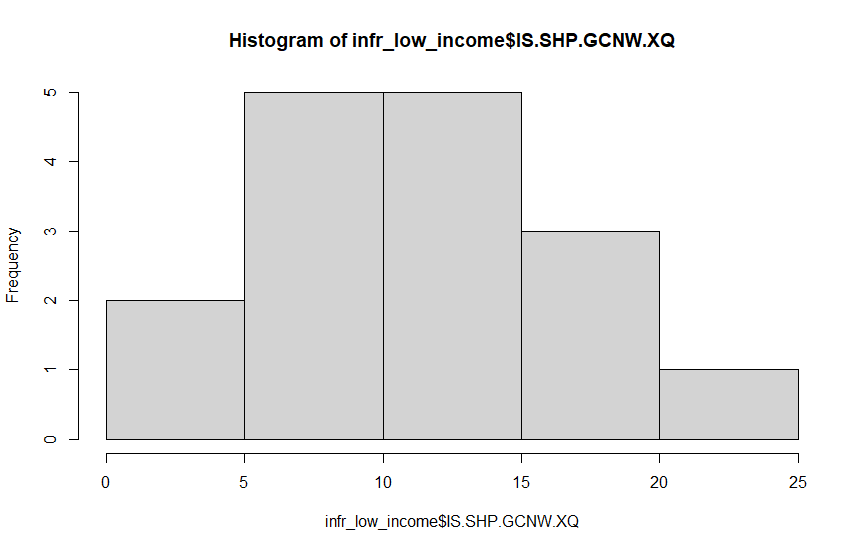
*hist(infr\_hight\_income$IS.SHP.GCNW.XQ)*

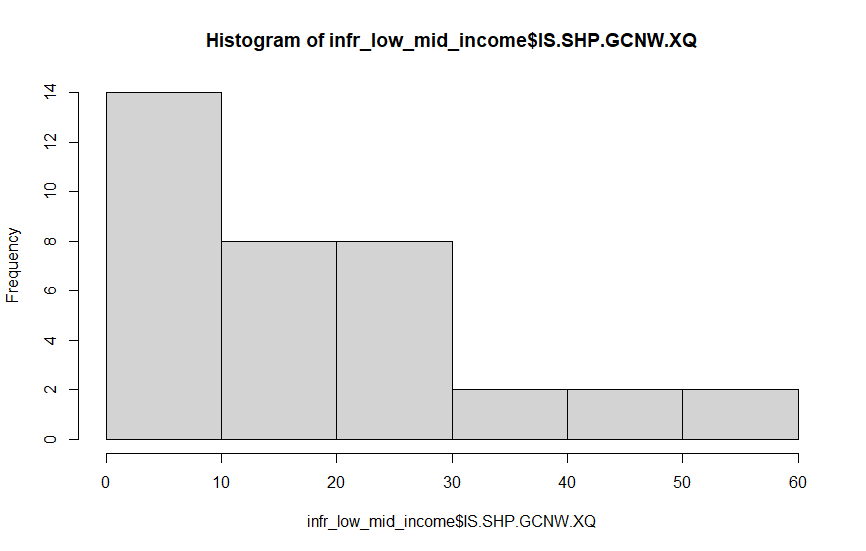
*hist(infr\_low\_income$IS.SHP.GCNW.XQ)*

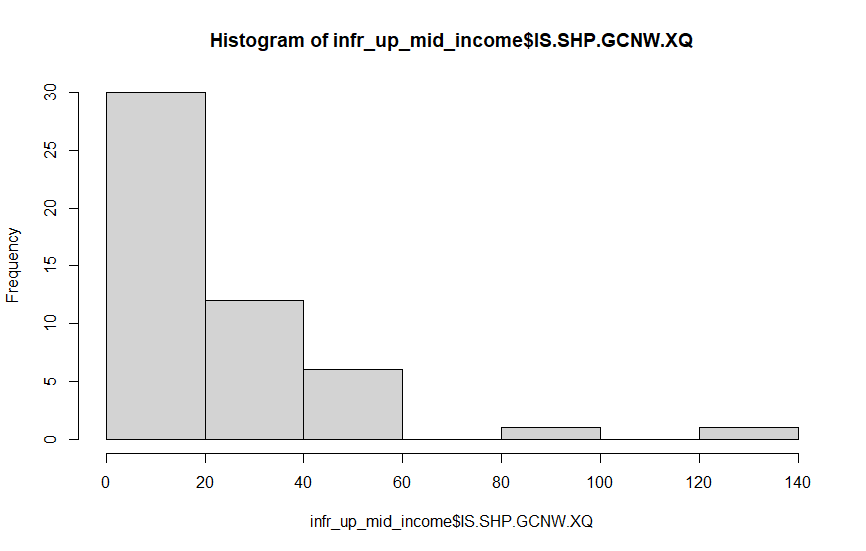
*hist(infr\_low\_mid\_income$IS.SHP.GCNW.XQ)*

*hist(infr\_up\_mid\_income$IS.SHP.GCNW.XQ)*









**Приведём описательную статистику каждой из выборок**

*summary(infr\_hight\_income$IS.SHP.GCNW.XQ)*

*summary(infr\_low\_income$IS.SHP.GCNW.XQ)*

*summary(infr\_low\_mid\_income$IS.SHP.GCNW.XQ)*

*summary(infr\_up\_mid\_income$IS.SHP.GCNW.XQ)*

A screenshot of a computer

Description automatically generated with medium confidence

**Визуализация по двум качественным и одному количественному (медианы) признакам**

*pd = position\_dodge(0.1)*

*ggplot(infr\_income, aes(x = income, y = IS.SHP.GCNW.XQ, color = region, group = region)) +*

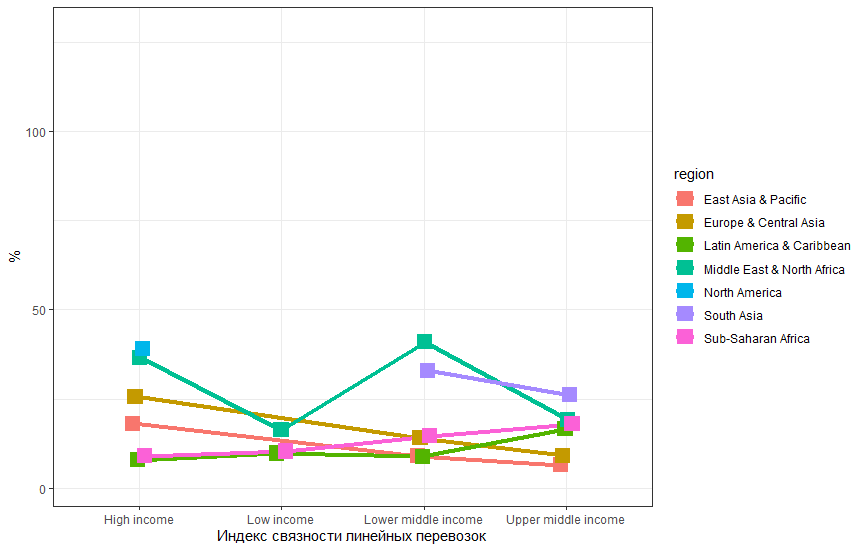
*stat\_summary(fun.data = median\_hilow, geom = 'line', size = 1.5, position = pd) +*

*stat\_summary(fun.data = median\_hilow, geom = 'point', size = 5, position = pd, pch = 15) +*

*theme\_bw() +*

*xlab('Индекс связности линейных перевозок') +*

*ylab('%')*

****

**Проверим гипотезу о принадлежности к нормальному закону распределения каждой выборки**

*shapiro.test(infr\_hight\_income$IS.SHP.GCNW.XQ)*

*shapiro.test(infr\_low\_income$IS.SHP.GCNW.XQ)*

*shapiro.test(infr\_low\_mid\_income$IS.SHP.GCNW.XQ)*

*shapiro.test(infr\_up\_mid\_income$IS.SHP.GCNW.XQ)*

Text

Description automatically generated

p-value < 0,05 – нулевая гипотеза отклоняется

Text

Description automatically generated with medium confidence

p-value > 0,05 – нулевая гипотеза принимается

Text

Description automatically generated

p-value < 0,05 – нулевая гипотеза отклоняется

Text

Description automatically generated

p-value < 0,05 – нулевая гипотеза отклоняется