Работа с графиками и диаграммами в R

3.1. Цель работы:

Научиться работать с графиками.

3.2. Общие сведения

R создает привлекательные диаграммы при минимальных затратах усилий с нашей стороны. Однако можно использовать графические параметры, чтобы назначать шрифты, цвета, типы линий, создавать оси, вспомогательные линии и аннотации. Подобная гибкость открывает перед вами широкие возможности оптимизации диаграмм. В этой работе мы начнем с простой диаграммы, а затем обсудим, как ее можно изменять и улучшать, чтобы диаграмма соответствовала вашим задачам. Затем мы рассмотрим усовершенствованные примеры, которые иллюстрируют более сложные методы изменения параметров диаграмм. Особое внимание будет уделено тем методам, которые могут быть использованы для самых разных типов диаграмм.

Для одновременного открытия нескольких окон графики и выбора необходимой диаграммы можно использовать функции *dev.new()*, *dev.next()*, *dev.prev()*, *dev.set()* и *dev.off()*. Эти команды работают под всеми операционными системами.

Пример 3.1:

1. Создадим заведомо функциональную зависимость:

```
> num<-c(1:10)
> square<-num*num
> plot(num,square,type="b")
```

Параметр **type="b"** означает, что на графике должны быть показаны *и точки, и линии*. Получившийся график показан на рис. 3.1.

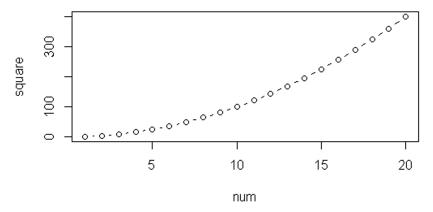


Рис.3.1. График квадратичной зависимости типа "b"

Если введете help(plot), чтобы узнать о других параметрах, то узнаете о следующих типах графиков:

- "p" for points,
- "1" for lines,
- "b" for both,

- "c" for the lines part alone of "b",
- "o" for both 'overplotted'.
- "h" for 'histogram' like (or 'high-density') vertical lines,
- "s" for stair steps,
- "S" for other steps, see 'Details' below,
- "n" for no plotting.

Попробуйте нарисовать различные типы графиков.

3.2. Общие сведения

Многие характеристики диаграмм (шрифты, цвета, оси, названия) можно изменять при помощи опций, которые называются «графические параметры».

Один способ назначить эти параметры – использовать функцию **par().** Значения параметров, заданные таким способом, будут действовать на протяжении всей сессии, пока вы не измените их. Формат применения функции таков:

```
par(название_параметра=значение, название_параметра=значение ,...).
```

Функция раг() без аргументов выводит на экран действующие значения графических параметров. Добавление аргумента no.readonly=TRUE позволяет увидеть только те графические параметры, которые можно изменять.

3.3. Как изменить параметры линий и точек

1. Изменяем ключевые точки на графике:

Продолжая наш пример, представим, что для обозначения отдельных точек вам захотелось использовать заполненный треугольник вместо пустого кружка и соединить символы пунктирной линией, а не сплошной. Это можно сделать при помощи следующего программного кода:

2. То же самое можно получить, указывая параметры прямо в функции plot():

Второй способ задать графические параметры — это включить записи типа название_параметра=значение внутри графической функции высокого уровня. В этом случае заданные параметры будут действовать только для конкретной диаграммы. Можно было бы построить тот же график при помощи следующего программного кода:

```
plot(num, square, type="b", lty=2, pch=17)
```

Таблица 3.1. Параметры для указания типов символов и линий

Параметр	Описание			
Pch	Определяет тип символа (см Рис. 3.2)			
Cex	Определяет размер символа. сех – это число, обозначающее, как символы должны быть			
	масштабированы по отношению к размеру по умолчанию. 1 = размер по умолчанию, 1.5			
	на 50% крупнее, 0.5 – на 50% мельче и т. д.			
Lty	Определяет тип линии (см Рис. 3.2)			
Lwd	Определяет толщину линии по сравнению с толщиной линии по			
	умолчанию (1). Например, lwd=2 делает линию в два раза толще,			
	чем по умолчанию			





Рис. 3.2. Параметры символов и типов линий для графиков.

3.3. Как работать с цветом

Таблица 3.2. Параметры для назначения цвета

Параметр	Описание
Col	Цвет элементов на графике. Для некоторых функций (таких как lines и ріе) можно
	указывать вектор из значений, которые используются по очереди. Например, если
	col=c ("red", "blue") и изображены три линии, первая будет красной, вторая - синей
	и третья – красной
col.axis	Цвет значений осей
col.lab	Цвет подписей осей
col.main	Цвет заголовков
col.sub	Цвет подзаголовков
Fg	Цвет графика
Bg	Цвет фона

В R цвета можно обозначать номером, названием, в шестнадцатеричной системе, а также в системах RBG или HSV. Например, col=1, col="white", col="#FFFFFF", col=rgb(1,1,1) и col=hsv(0,0,1) — взаимозаменяемые способы обозначить белый цвет. Функция rgb() определяет цвета по значениям красного, зеленого и синего, а hsv() основана на значениях оттенка и насыщенности.

В R также реализован ряд функций, которые позволяют создавать векторы из близких цветов. К таким функциям относятся rainbow(), heat.colors(), terrain.colors(), topo.colors() и cm.colors(). Например, rainbow(10) создает 10 соседних "радужных" цветов. Оттенки серого создаются функцией gray(). В этом случае вы задаете оттенки серого в виде вектора чисел от 0 до 1. Команда gray(0:10/10) создаст 10 оттенков серого. Попробуйте запустить программный код:

```
n <- 10
> mycolors <- rainbow(n)
> pie(rep(1, n), labels=mycolors, col=mycolors) # Круговая диаграмма
> mygrays <- gray(0:n/n)
> pie(rep(1, n), labels=mygrays, col=mygrays)
```

3.3. Характеристики текста

Таблица 3.3. Параметры, определяющие размер шрифта

Параметр	Описание
cex	Число, определяющее, как отображаемый на диаграмме текст будет масштабирован относительно размера по умолчанию (1). 1.5 - на 50% больше, 0.5 - на 50% меньше и т. д.

cex	axis	Размер значений на осях по отношению к сех		
cax	lab	Размер названий осей по отношению к сех		
cex	main	Размер заголовков по отношению к сех		
cex	sub	Размер подзаголовков по отношению к сех		

```
Hапример, после выполнения команды windowsFonts(
    A=windowsFont("Arial Black"),
    B=windowsFont("Bookman Old Style"),
    C=windowsFont("Comic Sans MS")
```

вы сможете использовать A, B и C как названия семейств шрифтов. В этом случае par(family="A") назначит шрифт Arial Black

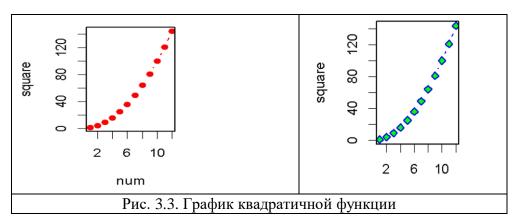
3.6. Размеры диаграммы и полей

Таблица 3.3. Параметры для определения размеров диаграммы и полей

Параметр	Описание	
Pin	Размер диаграммы (ширина, высота) в дюймах	
Mai	Числовой вектор, задающий размеры полей, где параметры с (низ, лево, верх, право) измеряются в дюймах	
Mar	Числовой вектор, задающий размеры полей, где параметры с (низ, лево, верх, право) измеряются в числе строк. По умолчанию это с (5, 4, 4, 2) +0.1	

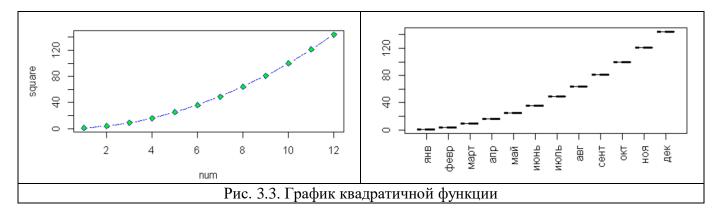
Давайте используем все параметры, которые мы успели обсудить, чтобы усовершенствовать наш простой пример. Представленный ниже программный код позволяет получить диаграммы, показанные на рис. 3.3.

```
opar <- par(no.readonly=TRUE)
par(pin=c(2, 3))
par(lwd=2, cex=1.5)
par(cex.axis=.75, font.axis=3)
plot(num, square, type="b", pch=19, lty=2, col="red")
plot(num, square, type="b", pch=23, lty=6, col="blue", bg="green")
par(opar)</pre>
```



Заменим цифры на оси х на название месяцев, для этого создадим и упорядочим вектор:

```
> month<-c("янв","февр","март" ,"апр" ,"май" ,"июнь" ,"июль","авг","сент",
"окт","ноя","дек")
> month.f=month
> month.o=ordered(month.f,levels=month)
> plot(month.o, square, type="b", pch=23, lty=6, col="blue", bg="green")
> plot(num, square, type="b", pch=23, lty=6, col="blue", bg="green", las=3)
```



Как видите, второй график, несмотря на идентичность параметров, стал дискретным. В команде изменился только разворот надписей (las=3).

Но при этом надо заметить, что использование нечисловых аргументов в графиках автоматически исключает возможности применения части настроек!

3.7. Добавление текста, настройка параметров осей и условных обозначений

Для многих графических функций высокого уровня (например, plot, hist, boxplot) возможен контроль не только графических параметров, но и параметров осей и надписей. К примеру, при помощи приведенного ниже программного кода можно разместить на диаграмме заголовок (main), подзаголовок (sub) и подписи осей (xlab, ylab), а также задать диапазон значений на осях (xlim, ylim).

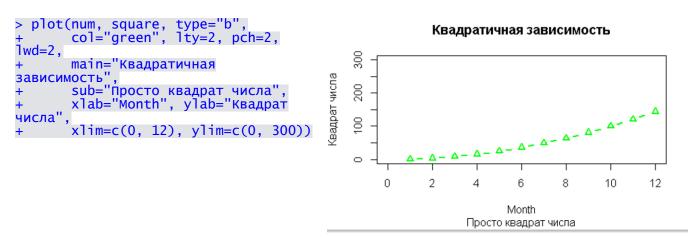


Рис. 3.3. Задаем заголовок для графика и подписи к осям.

Примечание. Некоторые графические функции высокого уровня по умолчанию выводят надписи и подписи на диаграммах. От них можно избавиться, указав ann=FALSE как один из аргументов команд plot() или par().

Для размещения заголовков и подписей осей на диаграмме можно также использовать функцию title(). Например:

```
> plot(num, square, type="b", ann=FALSE,
+ col="green", lty=2, pch=2, lwd=2,
+ xlim=c(0, 12), ylim=c(0, 300))
> title(main="квадратичная зависимость", col.main="red",
+ sub="Просто квадрат числа", col.sub="blue",
+ xlab="Month", ylab="квадрат числа",
+ col.lab="green", cex.lab=1)
Обязательно перечертить
график, иначе подписи
перекроют предыдущие.
```

Квадратичная зависимость

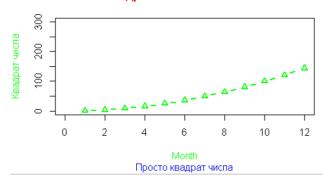


Рис. 3.6. Оформляем заголовок графика и подписи к осям.

Вместо осей, создаваемых на диаграммах по умолчанию, можно создать оси по своему усмотрению, используя функцию axis(). Формат ее применения таков (все параметры описаны в табл. 3.5):

axis(side, at=, labels=, pos=, lty=, col=, las=, tck=, ...).

Таблица 3.5. Параметры осей

Параметр	Описание
Side	Цифра, определяющая, с какой стороны диаграммы рисовать ось (1 = низ, 2 = лево, 3 = верх, 4 = право)
At	Числовой вектор, который задает положение делений на осях
Labels	Текстовый вектор, который содержит подписи под делениями осей (если вектор не задан, используются значения вектора а~)
Pos	Координата оси (то есть значение другой оси, в котором первая ось пересекает ее)
Ity	Тип линии
Col	Цвет линии и делений оси
Las	Положение подписей делений по отношению к оси (0 = параллельно, 2 = перпендикулярно, 3=вертикально)
Tck	Длина деления оси, выражается в виде доли от длины диаграмма (отрицательное число означает положение деления кнаружи от рамки диаграммы, положительное число - внутри рамки диаграммы, 0 - отсутствие делений, 1 - сетка); значение по умолчанию-0.01
()	Другие графические параметры

Когда вы сами создаете оси, нужно предотвратить появление осей, которые создаются по умолчанию графической функцией высокого уровня. Аргумент *axes=FALSE* подавляет создание всех осей (даже рамки вокруг диаграммы, если вы не добавите аргумент *frame.plot=TRUE*). Аргументы xaxt="n" и yaxt="n" отменяют создание x- и y-осей соответственно (при этом рамка без делений остается). Приведенный ниже программный код – это слегка избыточный надуманный пример, который демонстрирует применение всех аргументов, которые мы успели обсудить. График, который получается в результате, представлен на рис. 3.6.

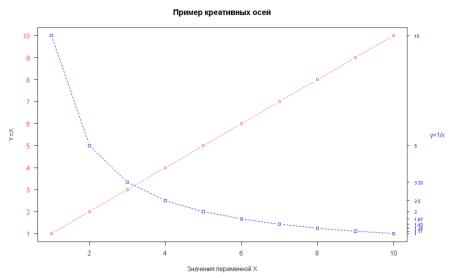


Рис. 3.7. Пример настройки параметров осей

Вам уже знакомы почти все команды из этого программного кода, за исключением lines() и mtext(). Команда plot() начинает построение новой диаграммы. Используя вместо этого команду lines(), вы можете добавить новые графические элементы к уже существующей диаграмме. Функция mtext() нужна для размещения текста на полях диаграммы.

Для полноты картины добавим легенду:

```
legend("topright", inset=.01, title="Два графика", c("Y=X","Y=1/x"),
+ lty=c(1, 2), pch=c(21, 22), col=c("red", "blue"))
```

Чтобы получить более подробную информацию о легендах, читайте *help(legend)*. Особенно информативны приведенные в файле справки примеры.

Таблица 3.6. Параметры легенды

Аргумент	Описание
Location	Существует несколько способов определить положение легенды на диаграмме. Можно указать х,у-координаты верхнего левого угла легенды. Можно использовать команду locator (1); в этом случае вы используете мышку, чтобы указать положение легенды. Также можно обозначать положение легенды при помощи ключевых слов bottom (внизу), bottonleft (внизу слева), left (слева), topleft (сверху слева), top (сверху), topright (сверху справа), right (справа), bottomright (внизу справа) или center (в центре). Если вы применяете одно из этих ключевых слов, можно также использовать параметр inset= для того, чтобы указать, как далеко (в долях от размера диаграммы) нужно сдвинуть легенду внутрь диаграммы
Title	Текстовая строка - название легенды (необязательно)

Legend	Текстовый вектор с расшифровкой значений легенды
	.Другие аргументы. Если легенда расшифровывает цвета линий, укажите col= и приведите вектор с названиями цветов. Если легенда поясняет типы символов, укажите pch- и приведите вектор с номерами символов. Если легенда указывает значения линий разной ширины или стиля, используйте Iwd- или ity- и вектор значений ширины или стиля линий. Для того чтобы включить в легенду закрашенные квадратики (часто применяется для столбчатых или круговых диаграмм), используйте fill=и вектор со значениями цветов
	у.intersp = 0.7 - с помощью этого параметра можно изменить межстрочное расстояние в легенде

И последнее в этом разделе - Аннотации

Текст можно добавлять на диаграмму при помощи команд text() и mtext(). Команда text() позволяет поместить текст внутри диаграммы, а функция mtext() размещает текст на одном из четырех полей диаграммы. Форматы применения таковы:

```
text(location, "Текст аннотации", pos, ...) mtext("Текст", side, line=n, ...)
```

3.8. Объединение диаграмм

В R очень легко объединить несколько диаграмм в одну общую, используя функции par() или layout(). Добавим графический параметр mfrow=c(nrows,ncols) в функцию par() для создания матрицы из диаграмм размером $nrows \times ncols$, которая будет заполнена по рядам. Для заполнения этой матрицы по столбцам нужно использовать параметр mfcol=c(nrows,ncols):

```
par(mfrow=c(2,2)) # будет четыре графика

plot(wt,mpg, main="Диаграмма рассеяния для \n расхода топлива и веса машины")

plot(wt,disp, main="Диаграмма рассеяния для \n объема двигателя и веса машины")

nist(wt, main="Pacnpeдeneниe значений \n веса машины")

par(opar)

detach(mtcars)

Parrpagnama рассения для рассения для рассения для объема двигателя и веса машины")

par(opar)

финтрамма рассения для рассения для рассения для объема двигателя и веса машины

Pacrpagnama рассения двигателя и веса машины

Pacrpagnama рассения двигателя и веса машины

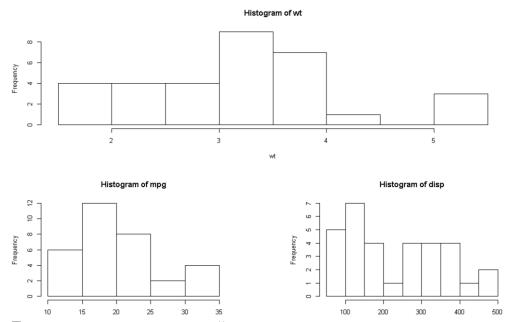
Pacrpagna
```

Рис. 3.8. Объединение графиков

Функция layout() имеет формат применения layout(mat), где mat — это матрица, в которой указаны положения нескольких совмещаемых диаграмм. При помощи приведенного программного кода одна диаграмма расположена в первом ряду, а две — во втором:

```
attach(mtcars)
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
```

hist(wt) hist(mpg) hist(disp) detach(mtcars)



Для указания размера каждой диаграммы можно использовать дополнительные аргументы функции layout(): widths= и heights=. Эти аргументы нужно применять в виде

widths= вектор значений ширины каждого столбца;

heights= вектор значений высоты каждой строчки.

```
attach(mtcars)
layout(matrix(c(1, 1, 2, 3), 2, 2, byrow = TRUE), widths=c(3, 1), heights=c(1, 2))
hist(wt)
hist(mpg)
hist(disp)
detach(mtcars)
```

3.9. Задания к лабораторной работе

В ходе лабораторной работы необходимо поработать с оформлением графиков и диаграмм. Необходимо учесть, что олимпийские игры по одному и тому же виду спорта бывают летние и зимние, соревнования делятся на мужские и женские.

- 1. Собрать данные согласно варианту по указанному виду и спорта в заданных странах.
- 2. Вывести графики динамики олимпийских достижений заданной страны по виду спорта относительно временной шкалы: источник (http://olympteka.ru/olymp/country/profile/rus.html) (Рис.3.9) столбчатую диаграмму по количеству мест 1-8 (спортсменов заданных стран) по каждой Олимпиаде по назначенному виду спорта, круговую диаграмму по количеству первых мест в каждой из олимпиад, функциональные графики тенденции изменения количества призовых мест отдельно по мужчинам и женщинам за последние 30 лет.
- 3. Вывести графики изменения спортивных достижений 1) по золотым медалям и 2) по призовым 3-местам по 7-и странам-призерам (разными цветами и точками) за последние 6 олимпиад. http://olympteka.ru/olymp/game/medals/51.html, например,

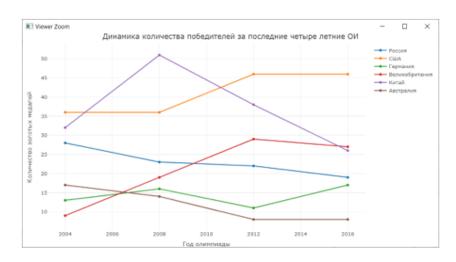


Рис 8. Динамика количества победителей за последние четыре летние ОИ (фигурное катание)

4. Используя тот же источник, выбрать информацию за последние **6** олимпиад по *заданному* виду спорта, аккумулировать данные по мужчинам и женщинам (отобразить на одном графике разными линиями), вывести динамику и статистику на графики (функциональный, столбчатый, круговой), используя изученные приемы оформления, в том числе массив графиков на одном поле. Пример функционального графика с пояснением:



Рис 9. Динамика количества призовых мест по фигурному катанию, занятых Россией, по годам (зимние ОИ) [второй gamacem)

Как видно из графика Россия по фигурному катанию каждый год занимала призовые места как минимум по одной дисциплине. В 2014 году Россия показала наилучшие результаты. Женщины обогнали мужчин, так среди мужчин не было призеров в индивидуальном зачете.

Обязательные требования к отчету: наличие исходных данных (фрагмент таблицы-скриншот из R и olympteka.ru), наличие подписей к осям графиков, заголовки к графику (main), масштабирование текста, наличие легенды, использовать объединение диаграмм и графиков (пункт 3.8), наличие трактовки каждого графика.

В качестве альтернативы можно скачать датасеты по летним и зимним олимпийским играм отсюда: https://www.kaggle.com/the-guardian/olympic-games и выполнить вышеназванные задания по ним, можно также использовать Википедию (по видам спорта, например, https://ru.wikipedia.org/wiki/Бадминтон_на_Олимпийских_играх)

Также обратите внимание на то, что необходимо отделять зимние и летние игры друг от друга — они хронологически перемешаны, но проходят в разных условиях.

Статистика по медалям и местам с 4-го по 8-ое

Место	Летние ОИ	•	(3)	0	Boe	4	5	6	7	8	
(4)	2016 РИО-ДЕ-ЖАНЕЙРО	19	18	19	56	11 4	165	6 ⁶	127	9 8	Призёры
	мужчины	9	6	12	27	8	7	5	8	5	
	женщины	10	12	7	29	3	9	1	4	4	
(4)	2012 ЛОНДОН	22	24	33	79	21 4	24 ⁵	13 ⁶	12 ⁷	18 ⁸	Призёры
	мужчины	11	9	17	37	10	14	6	7	8	
	женщины	11	15	16	42	11	10	7	5	10	
(3)	2008 ПЕКИН	23	21	29	73	20 4	235	19 6	167	17 ⁸	Призёры
	мужчины	12	8	21	41	7	12	7	8	11	
	женщины	11	13	8	32	13	11	11	8	6	
	микст	0	0	0	0	0	0	1	0	0	
(3)	2004 АФИНЫ	28	26	36	90	19 4	18 ⁵	11 ⁶	12 ⁷	13 ⁸	Призёры
	мужчины	16	10	25	51	8	10	7	7	3	
	женщины	12	16	11	39	11	8	4	5	10	
(2)	2000 СИДНЕЙ	32	28	29	89	20 4	10 ⁵	20 €	217	13 ⁸	Призёры
	мужчины	20	13	20	53	9	4	8	13	7	
	женщины	12	15	9	36	11	6	11	8	6	
	микст	0	0	0	0	0	0	1	0	0	
(2)	1996 АТЛАНТА	26	21	16	63	19 4	30 ⁵	15 ⁶	15 ⁷	11 ⁸	Призёры

Рис .3.9. Статистика по медалям спортсменов Росии (фрагмент)- выбрать для своей страны и своего вида спорта.

Варианты:

Baj	варианты:						
No	ФИО	Страна	Вид спорта				
1.	Алмаева Анастасия Ильинична	Россия	Фигурное катание				
2.	Блягоз Амаль Хазретович	Германия	Плавание				
3.	Вавакин Владислав Олегович	CIIIA	Баскетбол				
4.	Воробьев Артем Олегович	Китай	Гимнастика спортивная				
5.	Гаджиев Ахмед Русланович	Индонезия	Бадминтон				
6.	Задикян Аветис Арутюнович	Россия	Велоспорт				
7.	Иванченко Павла Андреевна	Турция	Тяжелая атлетика				
8.	Королев Сергей Юрьевич	Россия	Скелетон				
9.	Лотарев Сергей Юрьевич	США	Конькобежный спорт				
10.	Мазуренко Артём Алексеевич	Италия	Фехтование				
11.	Матюха Филипп Андреевич	Австралия	Плавание				
12.	Пономарь Дарья Сергеевна	Норвегия	Парусный спорт				
13.	Сидоренко Александр Андр	Норвегия	Биатлон				
14.	Смирнов Никита Олегович	Россия	Синхронное плаванье				
15.	Стройный Александр Алекс	Испания	Вольная борьба				
16.	Чеуж Асиет Асланбиевна	Германия	Спортивная гимнастика				
17.	Агаджанян Алёна Самвеловна	Канада	Хоккей на льду				
18.	Аникин Марк Андреевич	Швейцария	Горнолыжный спорт				
19.	Бачурин Иван Алексеевич	Румыния	Гребля академическая				
20.	Борисов Никита Алексеев	Нидерланды	Конькобежный спорт				
21.	Воробьев Игорь Александр	Россия	Борьба				
22.	Гаранина Людмила Виталь	Германия	Фигурное катание				
23.	Григоренко Никита Алексе	Великобритания	Футбол				
24.	Искрич Александр Алексан	Россия	Гимнастика художественная				
25.	Косенко Данил Сергеевич	Австрия	Горнолыжный спорт				
26.	Масенко Мария Сергеевна	Швеция	Лыжные гонки				
27.	Миков Никита Сергеевич	Россия	Прыжки в воду				
28.	Небывалов Максим Алекса	Бразилия	Футбол				
29.	Пикулев Андрей Сергеевич	Норвегия	Фристайл				
30.	Решетка Даниил Владим	Германия	Велоспорт				
31.	Сгонник Николай Сергее	Россия	Спортивная гимнастика				
32.	Таран Владимир Сергеевич	Китай	Теннис настольный				

№	ФИО	Страна	Вид спорта
1.	Абраамян Кристина Гургеновна	Норвегия	Биатлон
2.	Барсука Жамал (бюджет)	Австралия	Плавание
3.	Дзамихов Залим Виктор	США	Баскетбол
4.	Казарян Вероника Григор	Великобритания	Бадминтон
5.	Кличенко Давид Артурович	Россия	Гимнастика спортивная
6.	Козлов Эдуард Дмитриевич	Германия	Велоспорт
7.	Кочнев Владимир Юрьевич	США	Парусный спорт
8.	Крапоткин Максим Василь	Швейцария	Горные лыжи
9.	Краснослободцева Екатерина Олеговна	Корея	Геннис
10.	Синьков Дмитрий Вадим	Италия	Велоспорт
11.	Сластенов Константин Вяче	Россия	Плавание
12.	Усков Артём Алексеевич	Америка	Легкая атлетика
13.	Цобехия Данил Темурович	Франция	Велоспорт
14.	Шаблин Никита Дмитриев	Турция	Парусный спорт
15.	Арутюнян Артур Камоевич	Южная Корея	Стрельба из лука
16.	Аругюнян Рафаэль Гарегин	США	Теннис
17.	Белокобыльский Богдан Витал	Россия	Легкая атлетика
18.	Вебер Артем-Дариус Алексе	Китай	Геннис
19.	Волков Андрей Александ	Венгрия	Гребля на байдарках и каноэ
20.	Дерябин Андрей Виктор	Нидерланды	Конькобежный спорт
21.	Дудо Сергей Николаевич	Куба	Бокс
22.	Дука Виталий Андреевич	Россия	Беговые лыжи
23.	Журавлёв Даниил Дмитриев	США	Велоспорт
24.	Логвина Анна Владимир	Турция	Борьба Греко-римская
25.	Маркелов Владислав Андреев	Австралия	Конный спорт
26.	Осипов Василий Романович	Швеция	Фигурное катание
27.	Пшеничнов Андрей Алекс	Китай	Легкая атлетика
28.	Чертоусов Владимир Сергеев	Белоруссия	Биатлон