

КОРРЕЛЯЦИЯ ПИРСОНА

КОРРЕЛЯЦИЯ ПИРСОНА

- » Коэффициент корреляции Пирсона — мера силы линейной взаимосвязи:

$$r_{X_1 X_2} = \frac{\mathbb{E}((X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2))}{\sqrt{\mathbb{D}X_1 \mathbb{D}X_2}}$$

$$r_{X_1 X_2} \in [-1, 1]$$

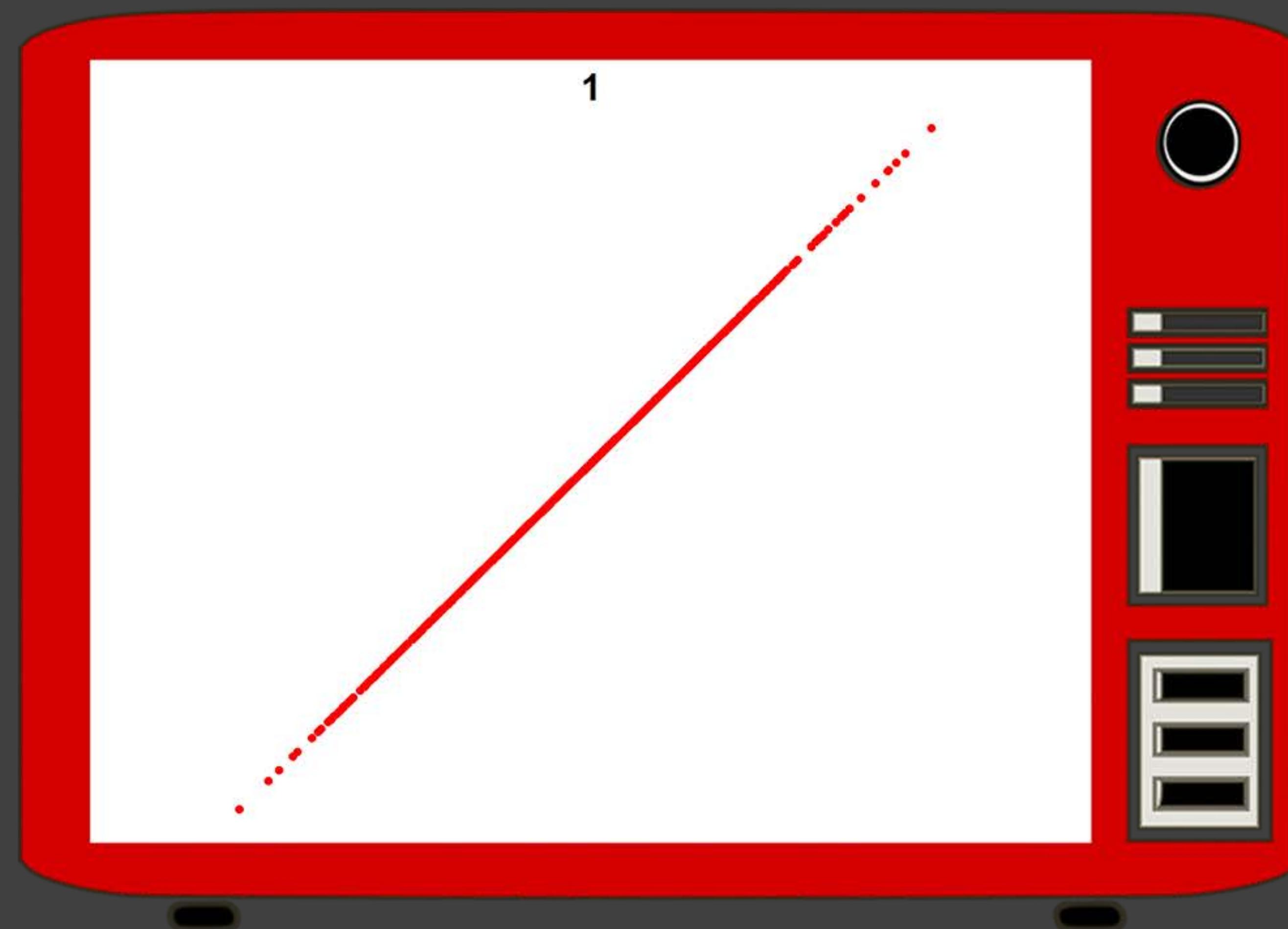
ВЫБОРОЧНЫЙ КОЭФФИЦИЕНТ

- » Выборка пар $(X_{1i}, X_{2i}), i = 1, \dots, n$
- » Выборочный коэффициент корреляции Пирсона:

$$r_{X_1 X_2} = \frac{\sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sqrt{\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 \sum_{i=1}^n (X_{2i} - \bar{X}_2)^2}}$$

СЖИМАЕМ ОБЛАКО





СГИБАЕМ ОБЛАКО



СЛОЖНЫЕ ЗАВИСИМОСТИ

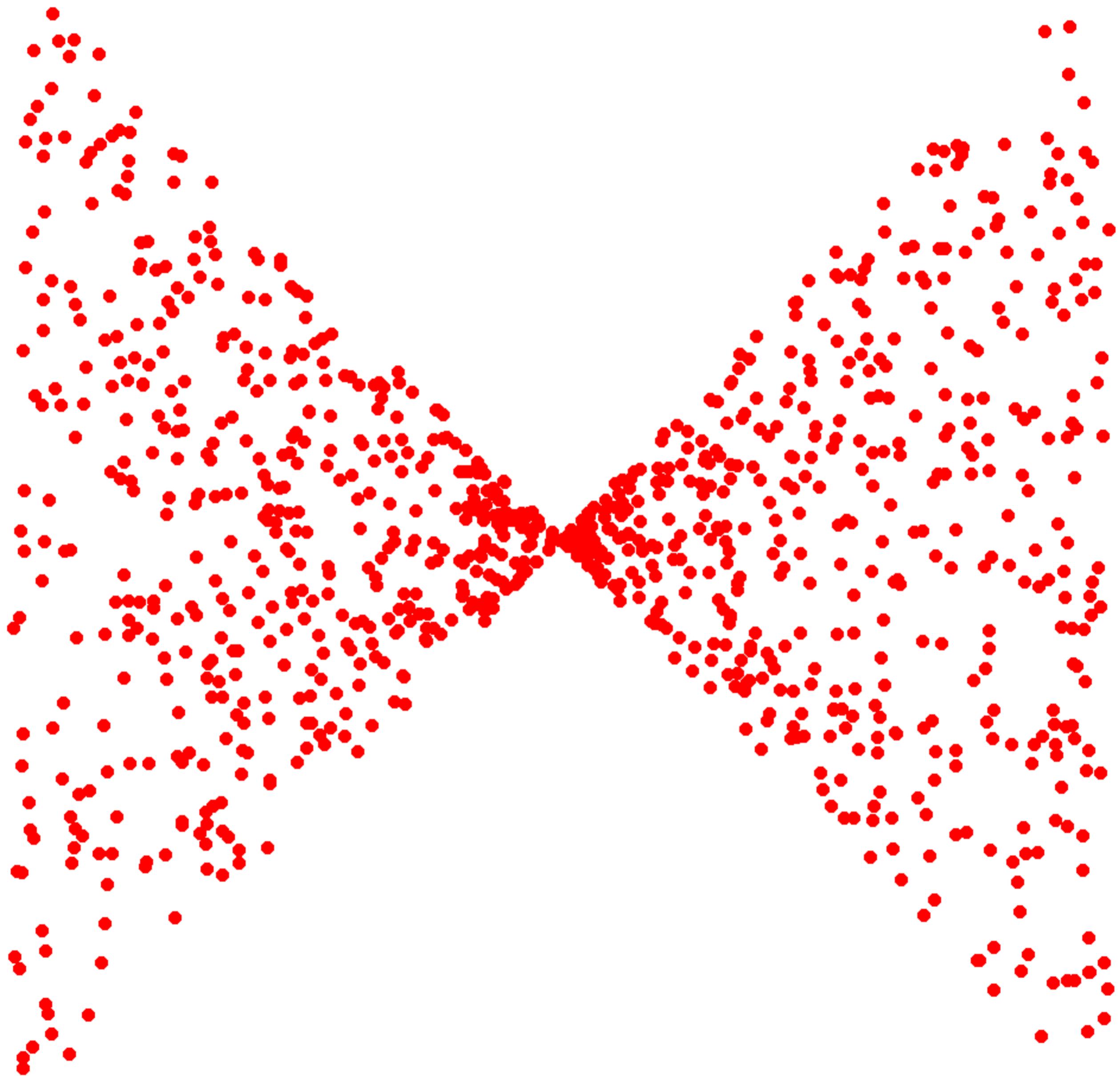


ВЫБРОСЫ

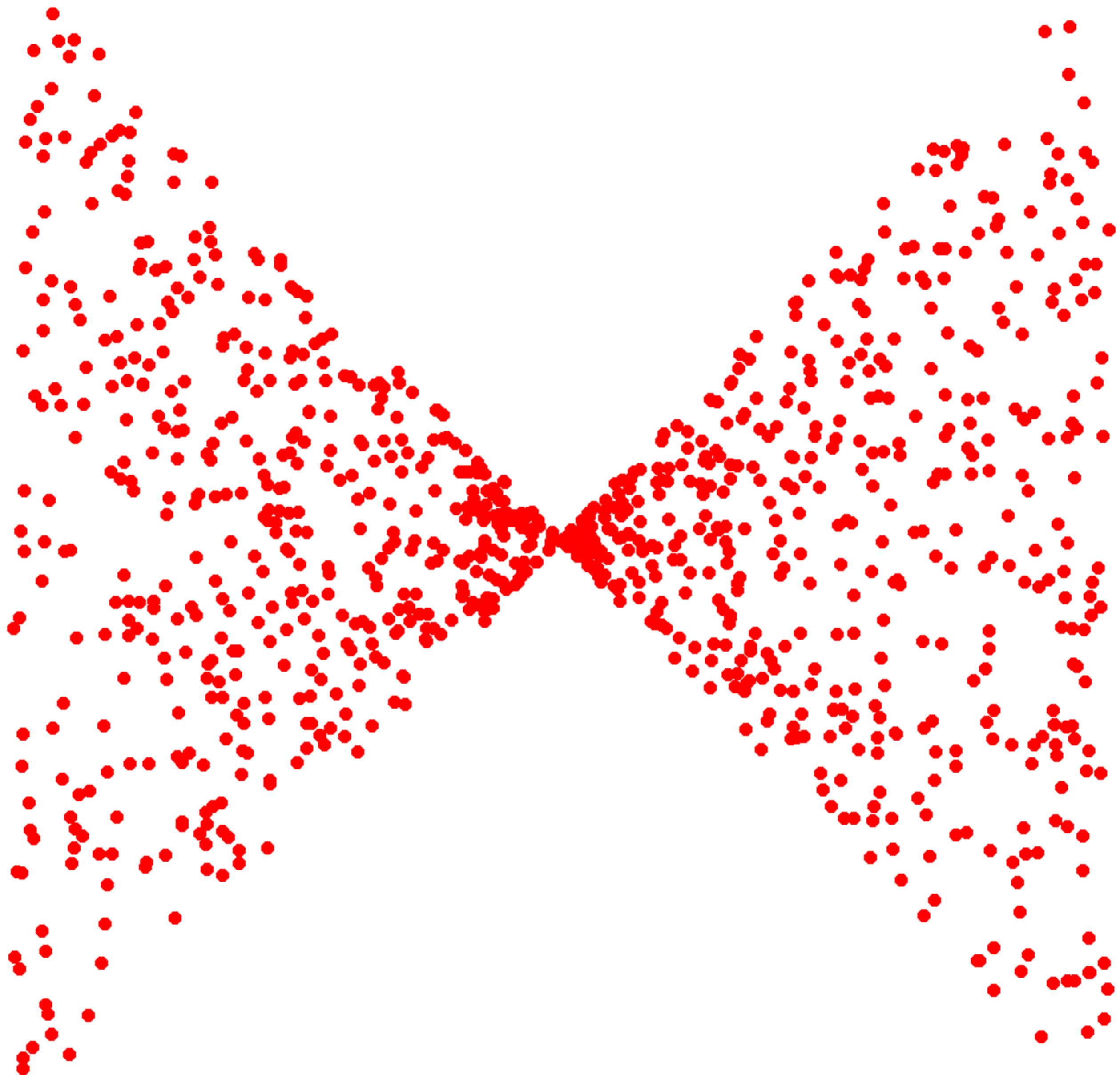


- › Коэффициент корреляции Пирсона — мера силы линейной взаимосвязи
- › Далее: корреляция Спирмена

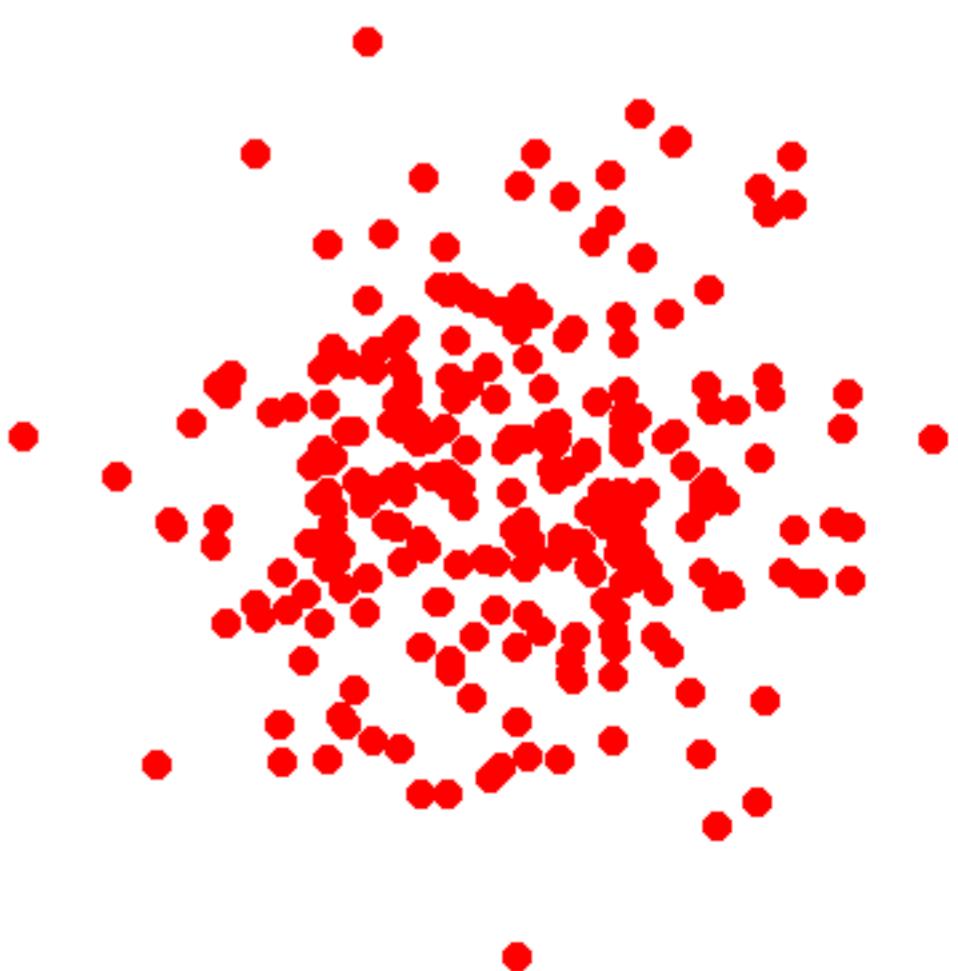
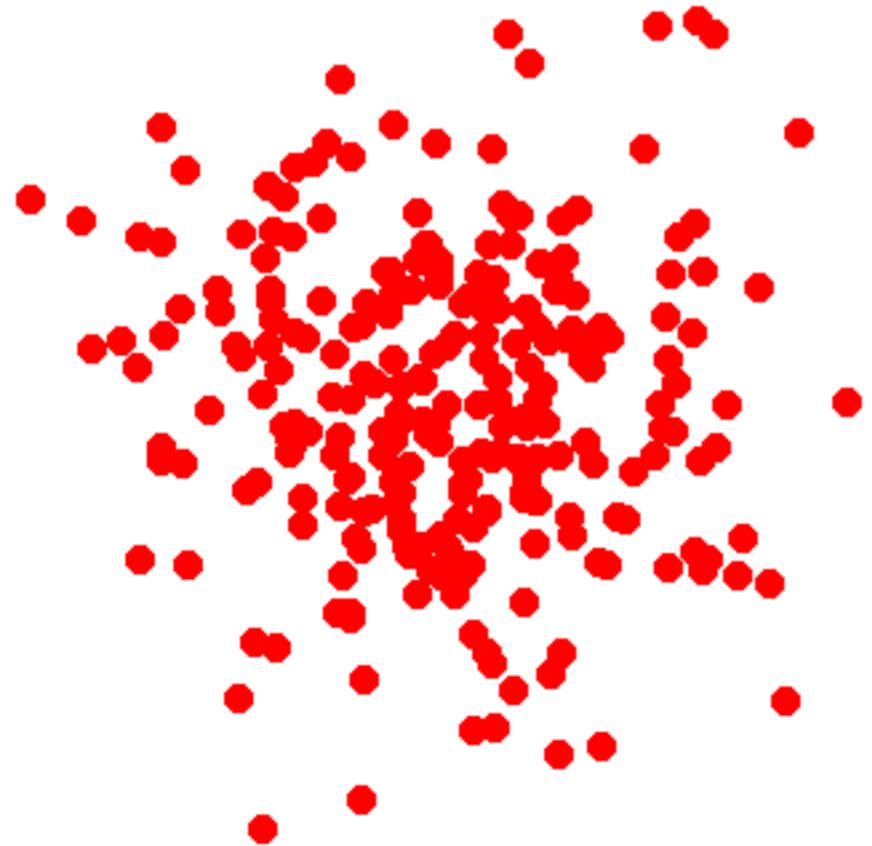
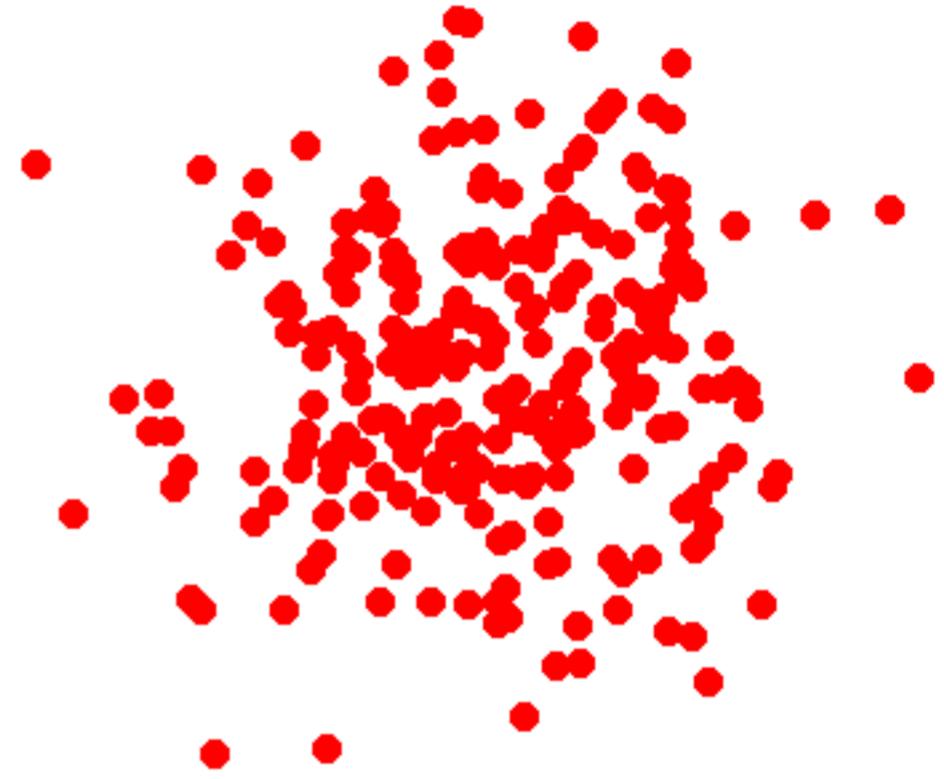
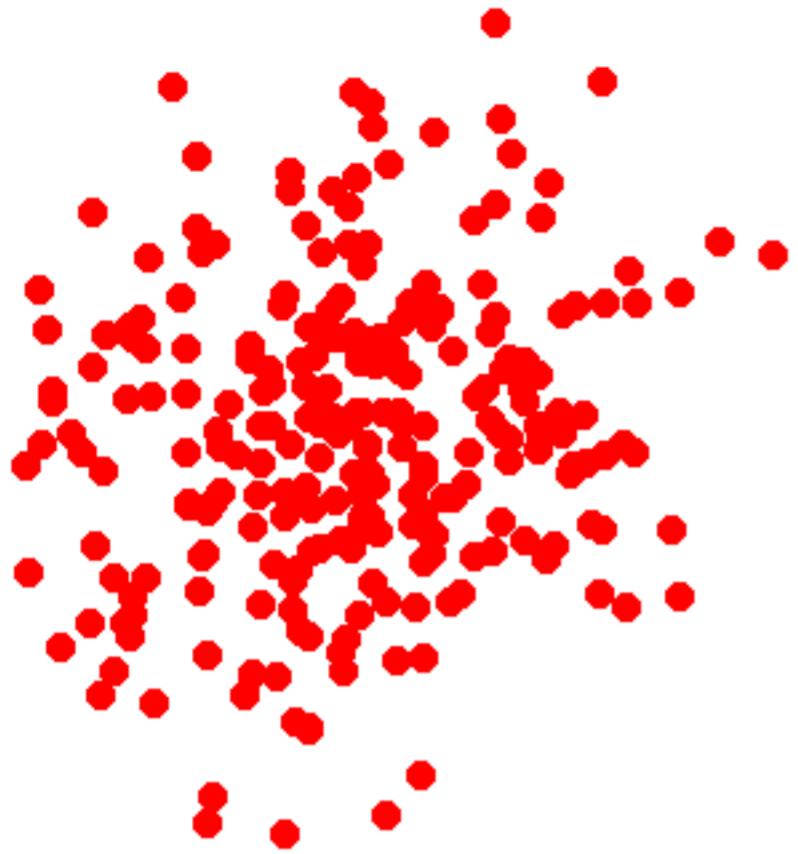
-0.0169



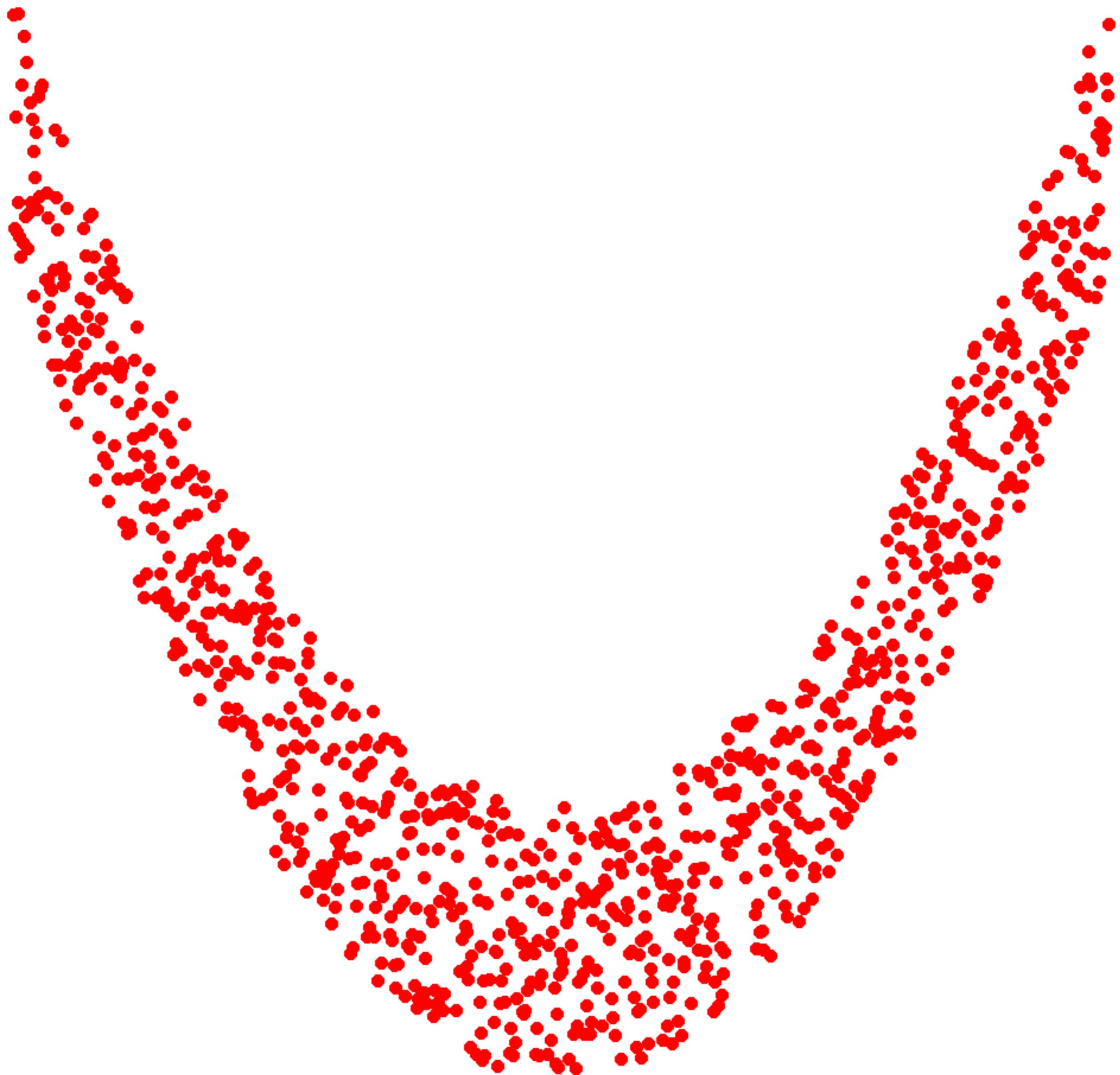
-0.0169



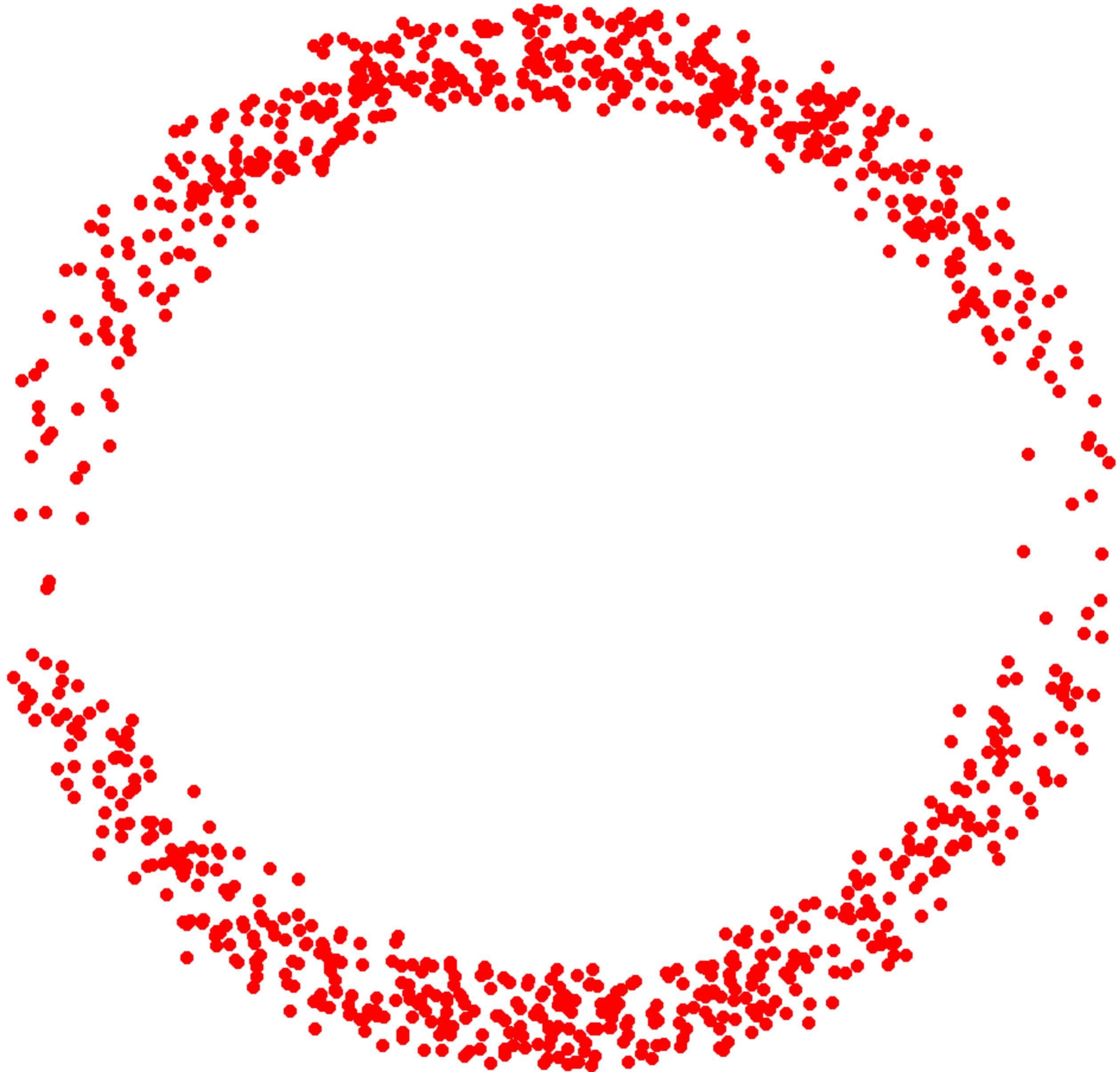
0.01105



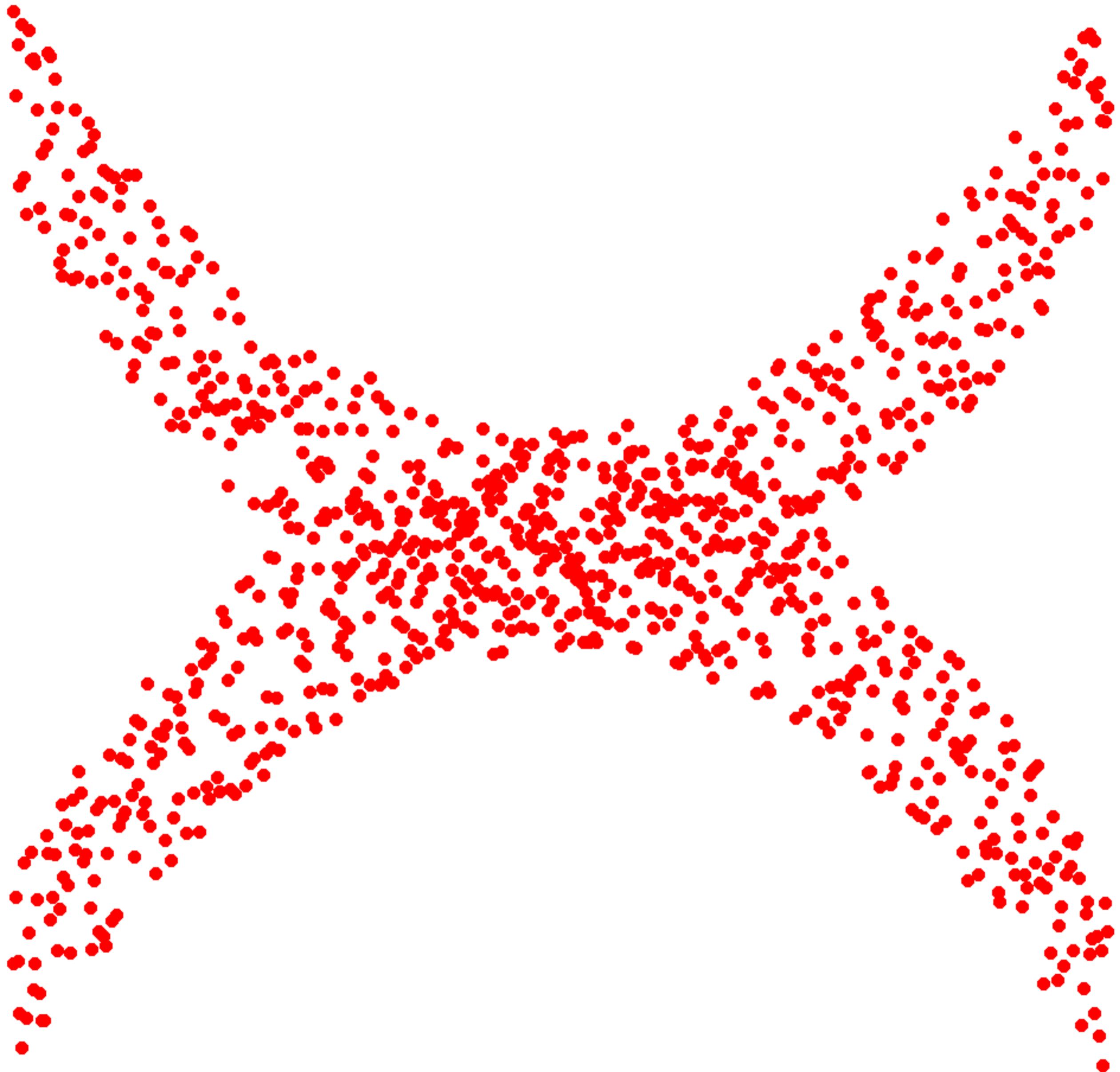
-0.00651



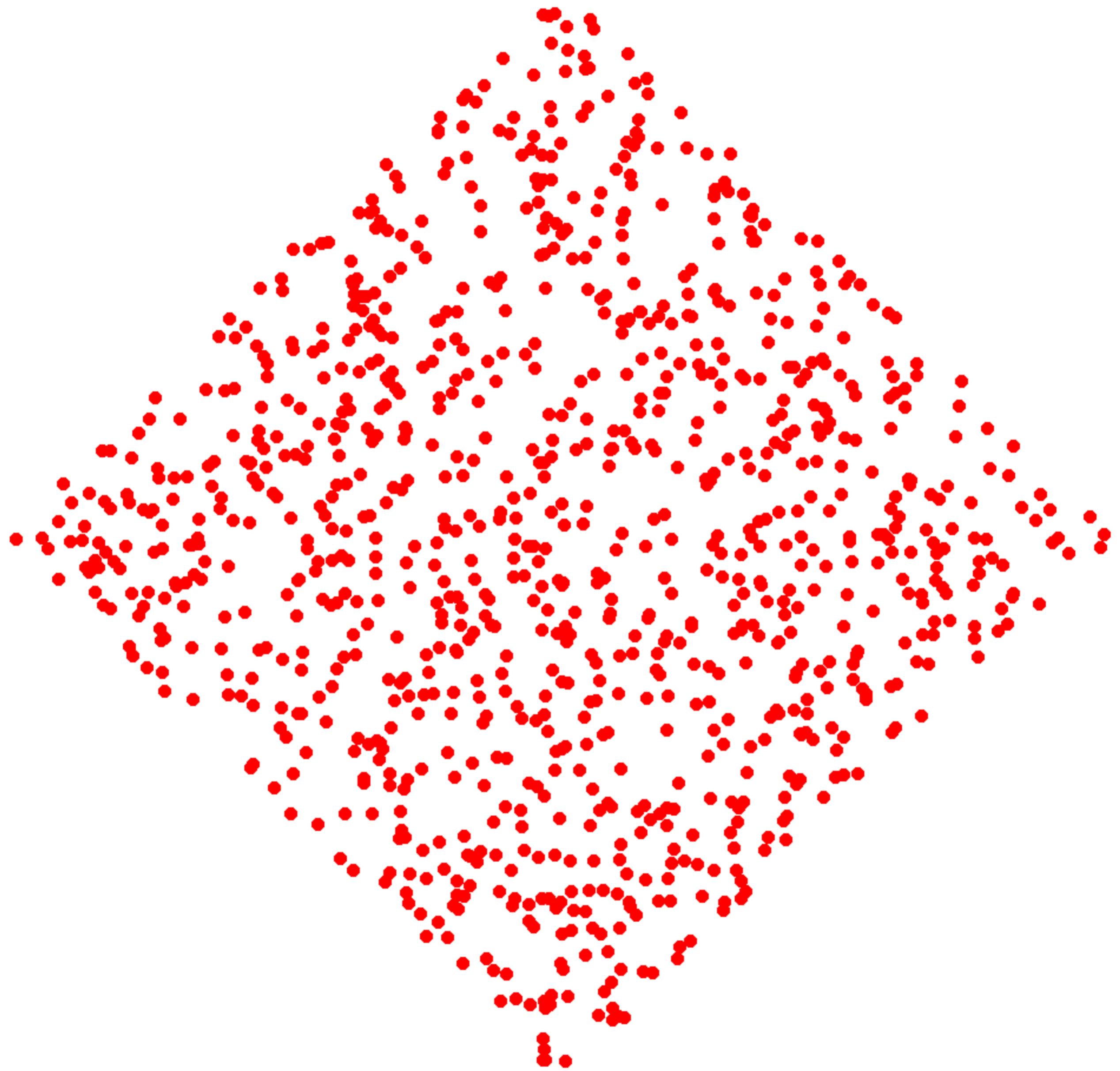
0.01102



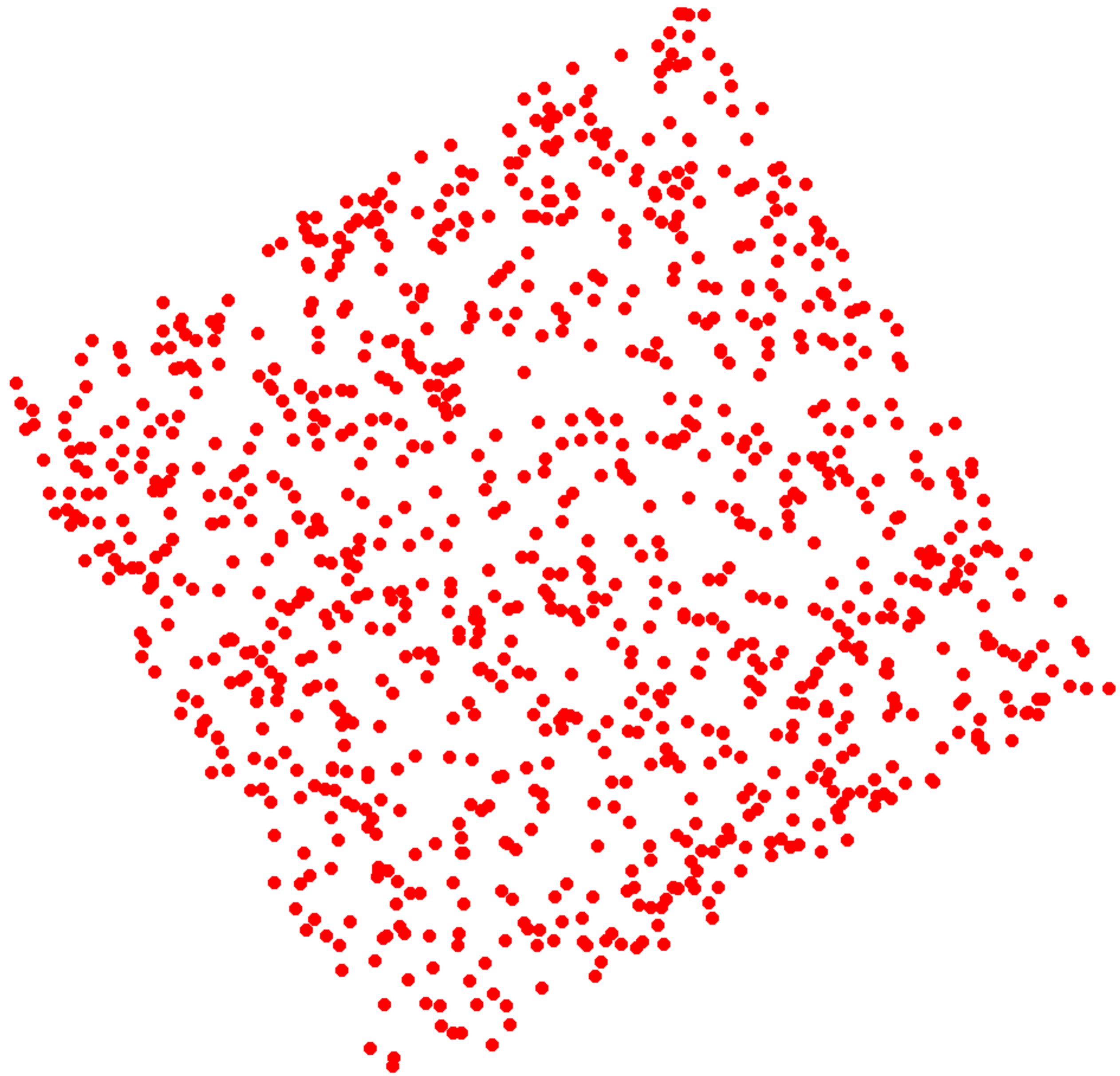
0.00538



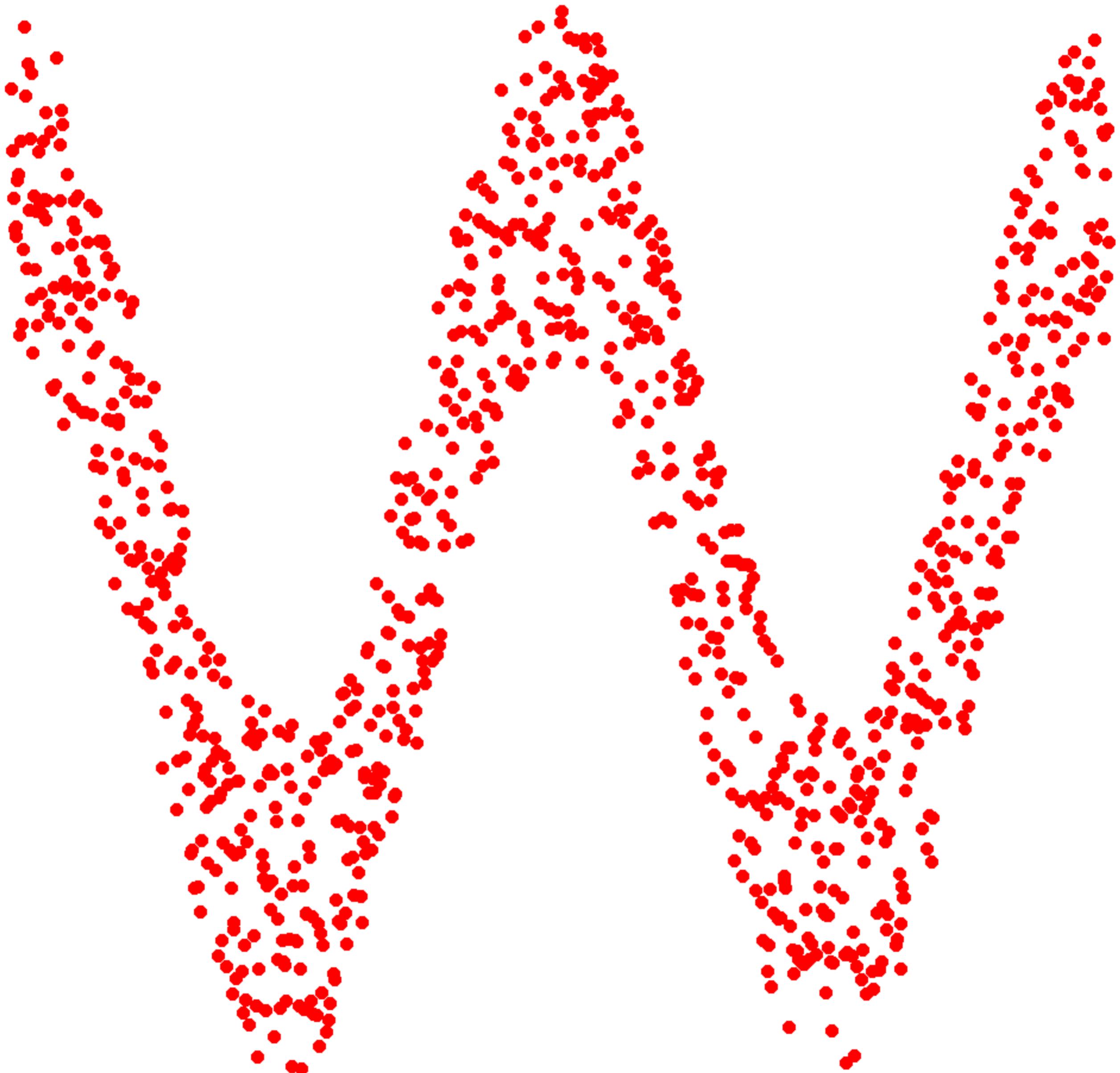
-0.01352



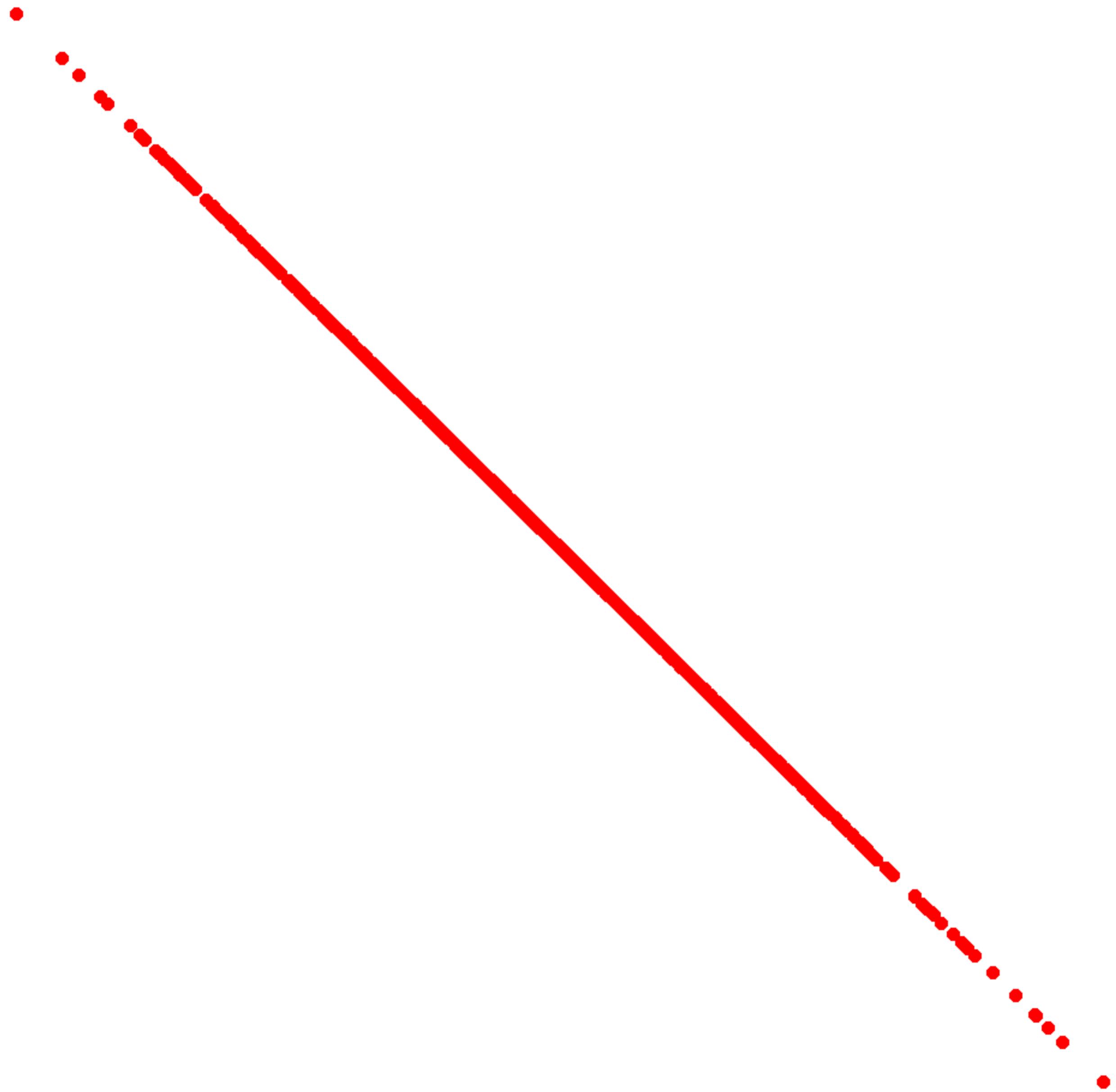
-0.03648



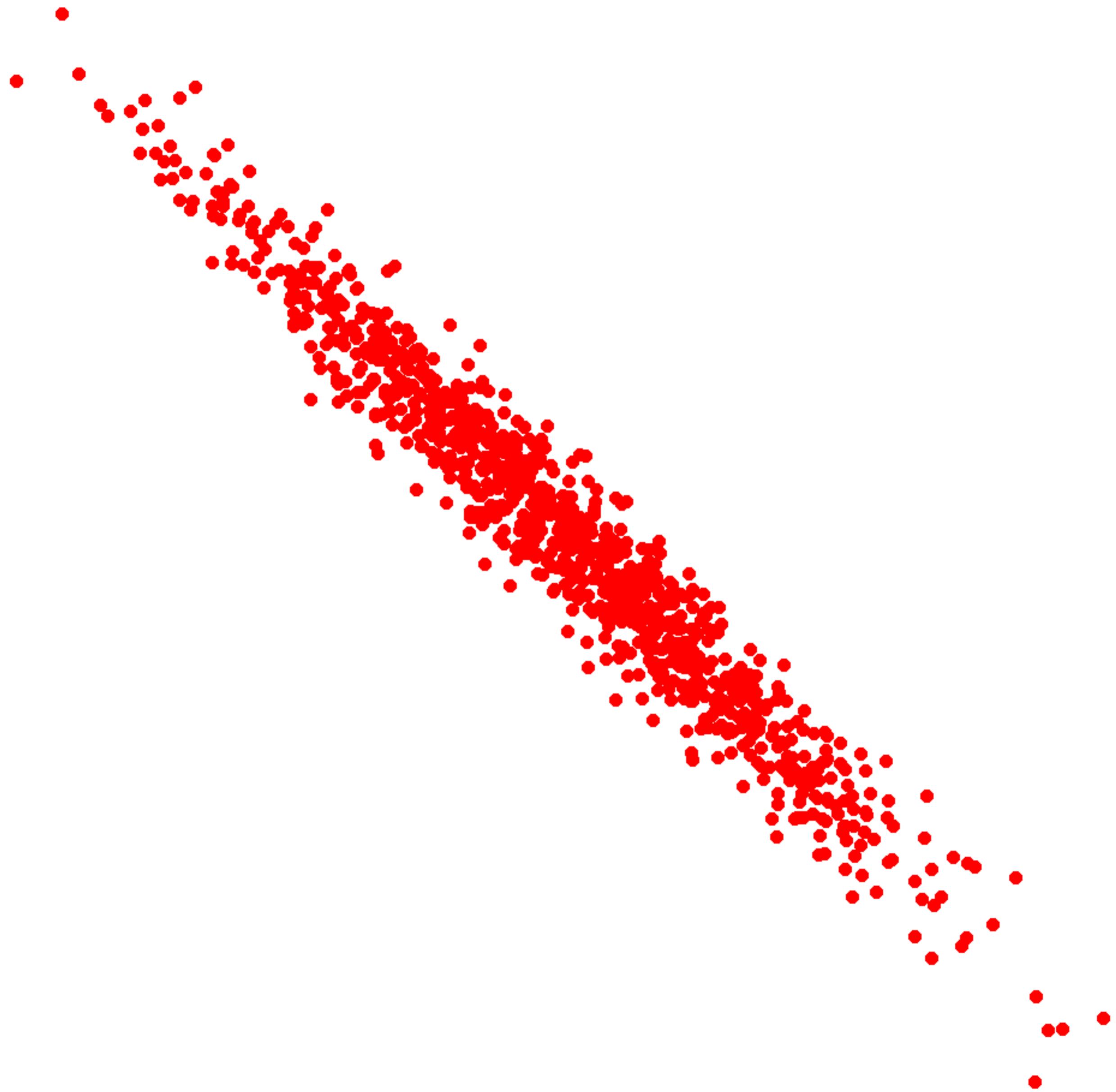
0.02051



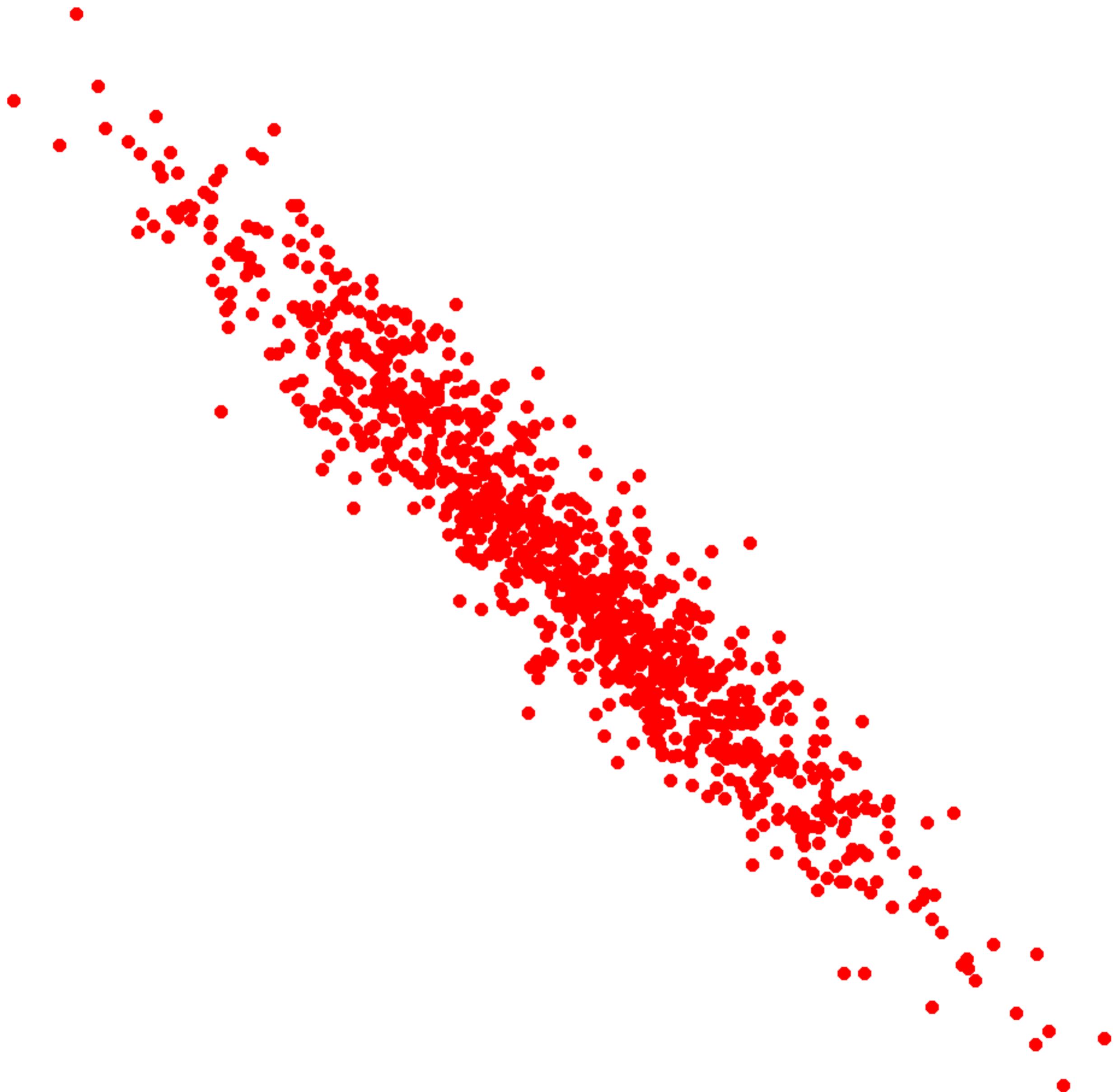
-1



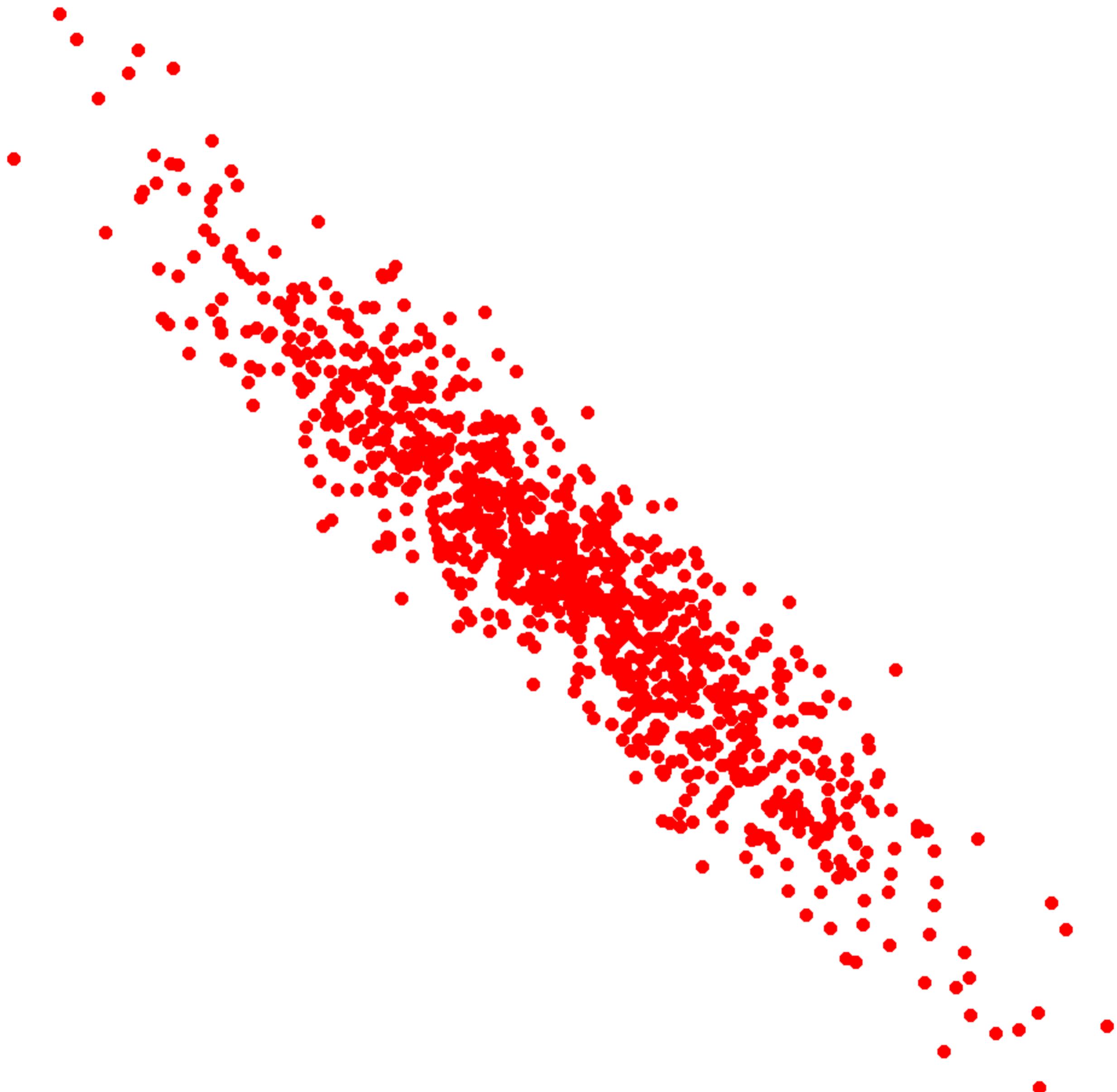
-0.97



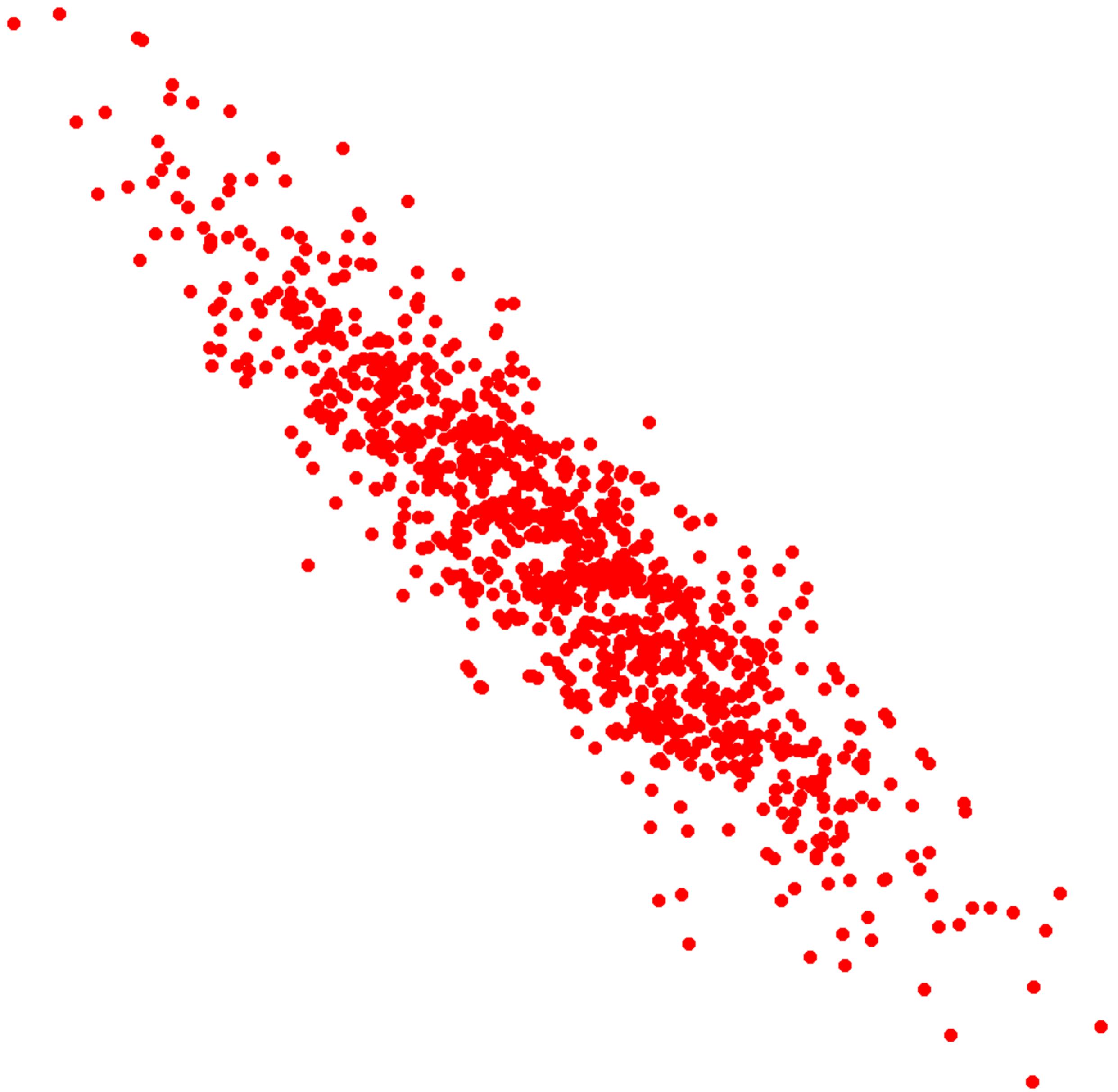
-0.95



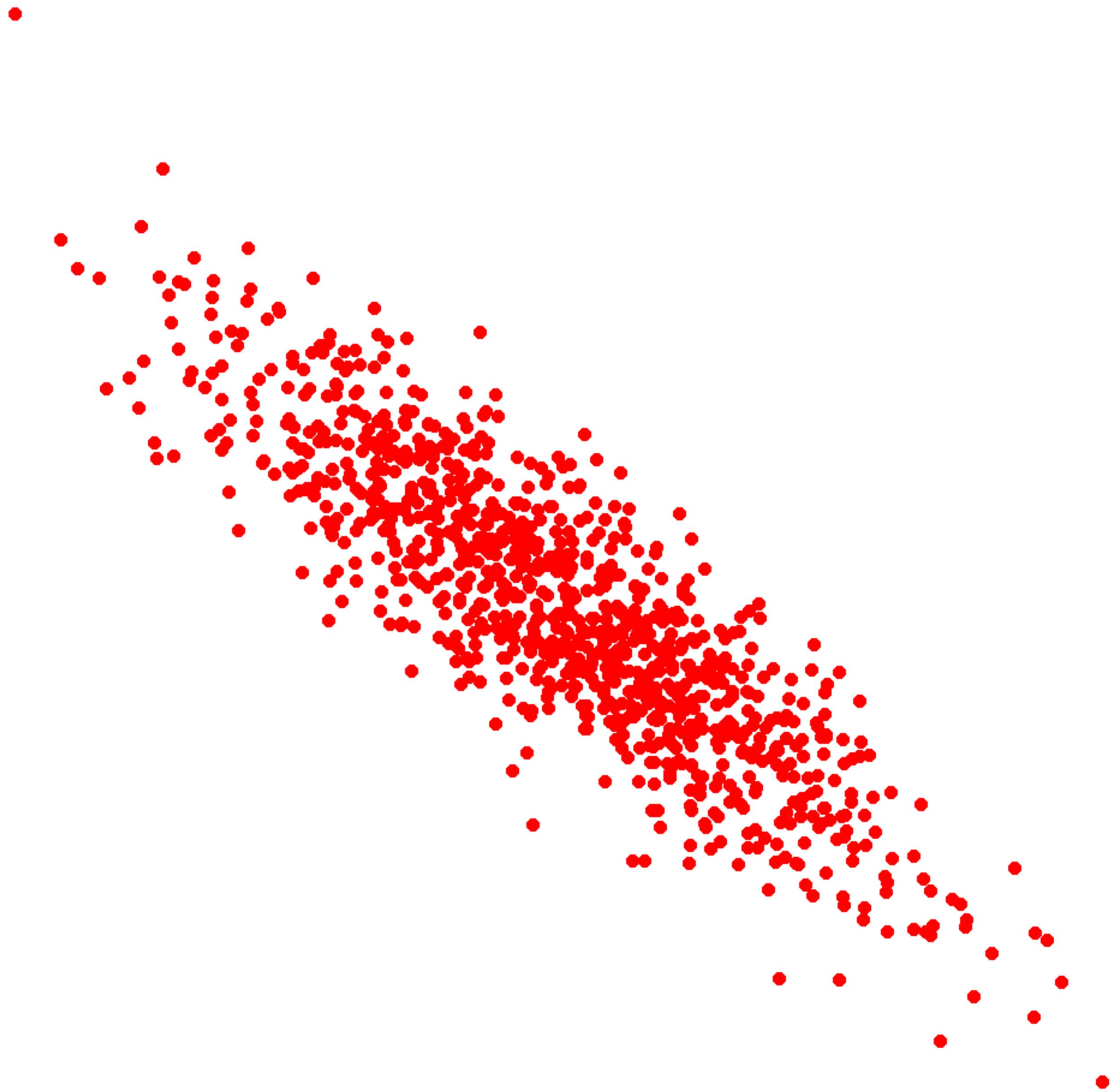
-0.92



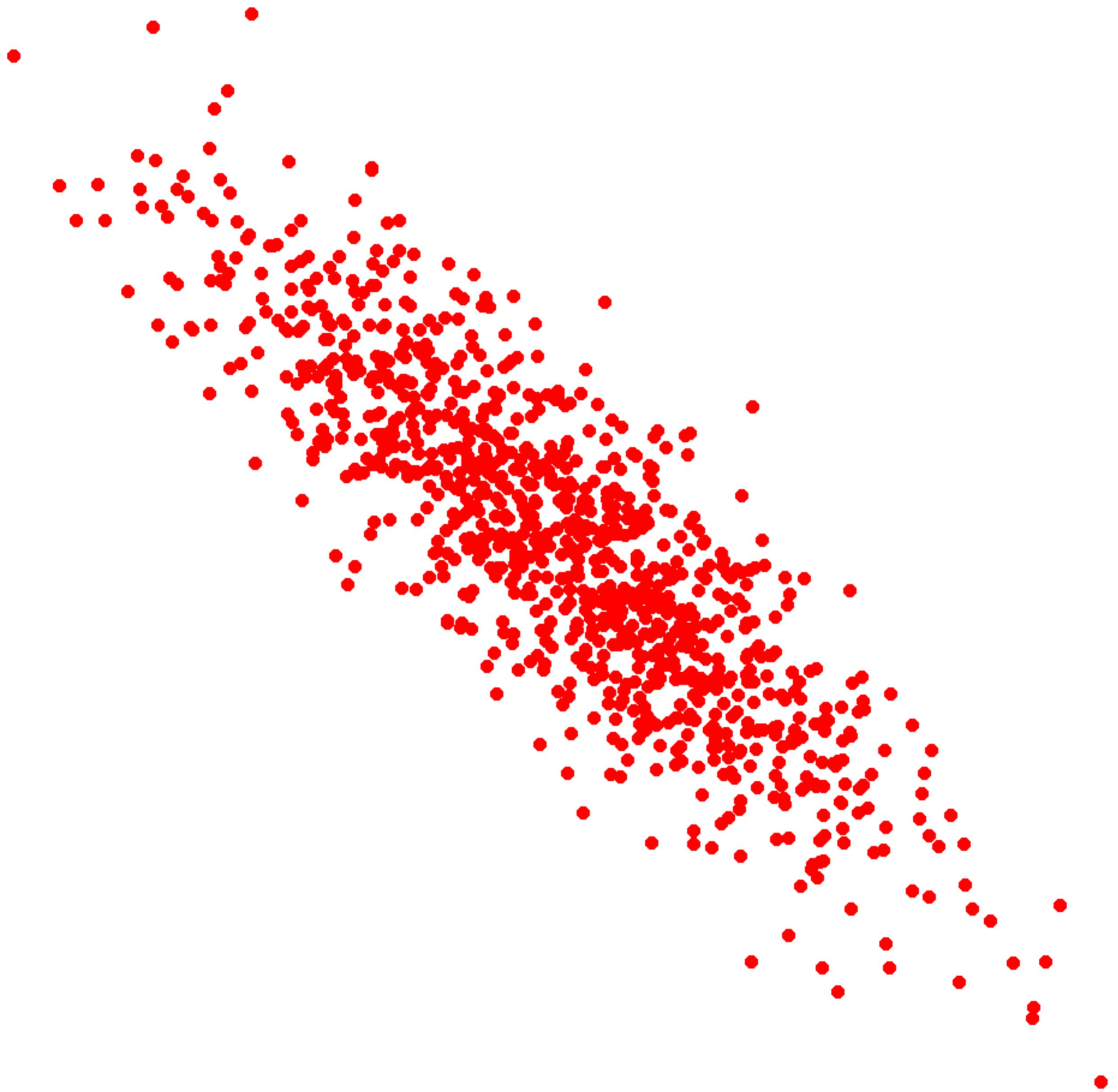
-0.9



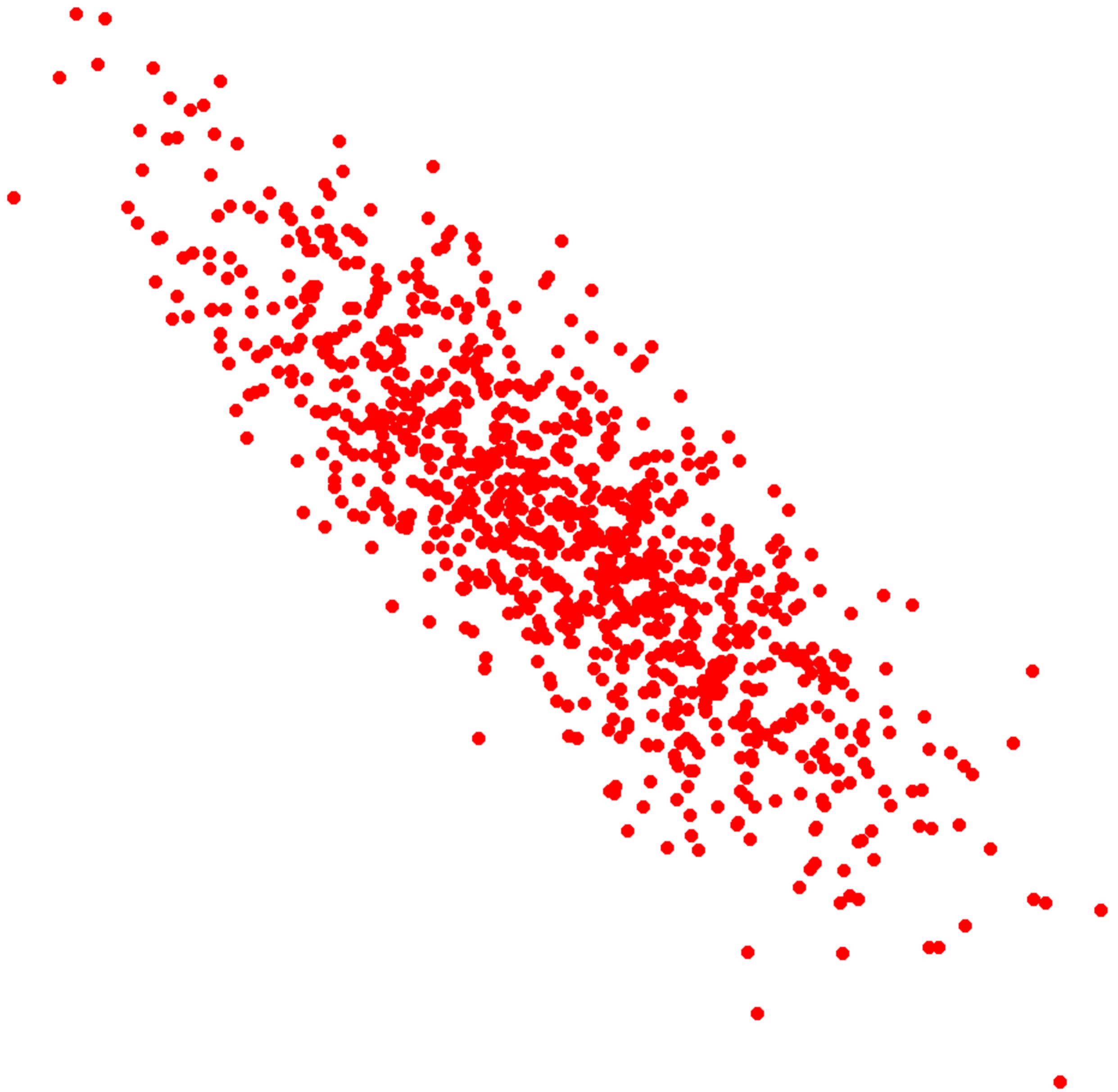
-0.87



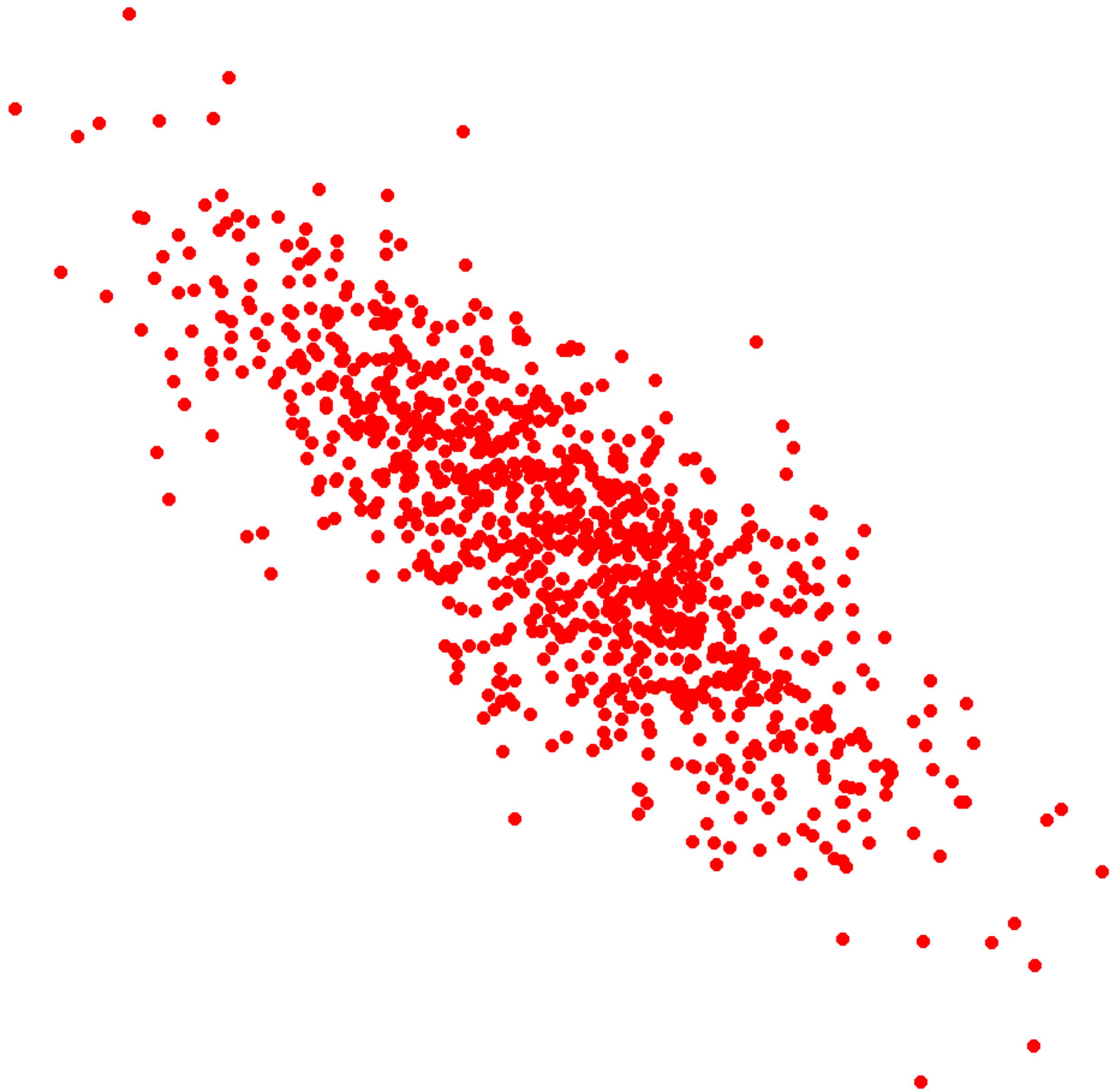
-0.86



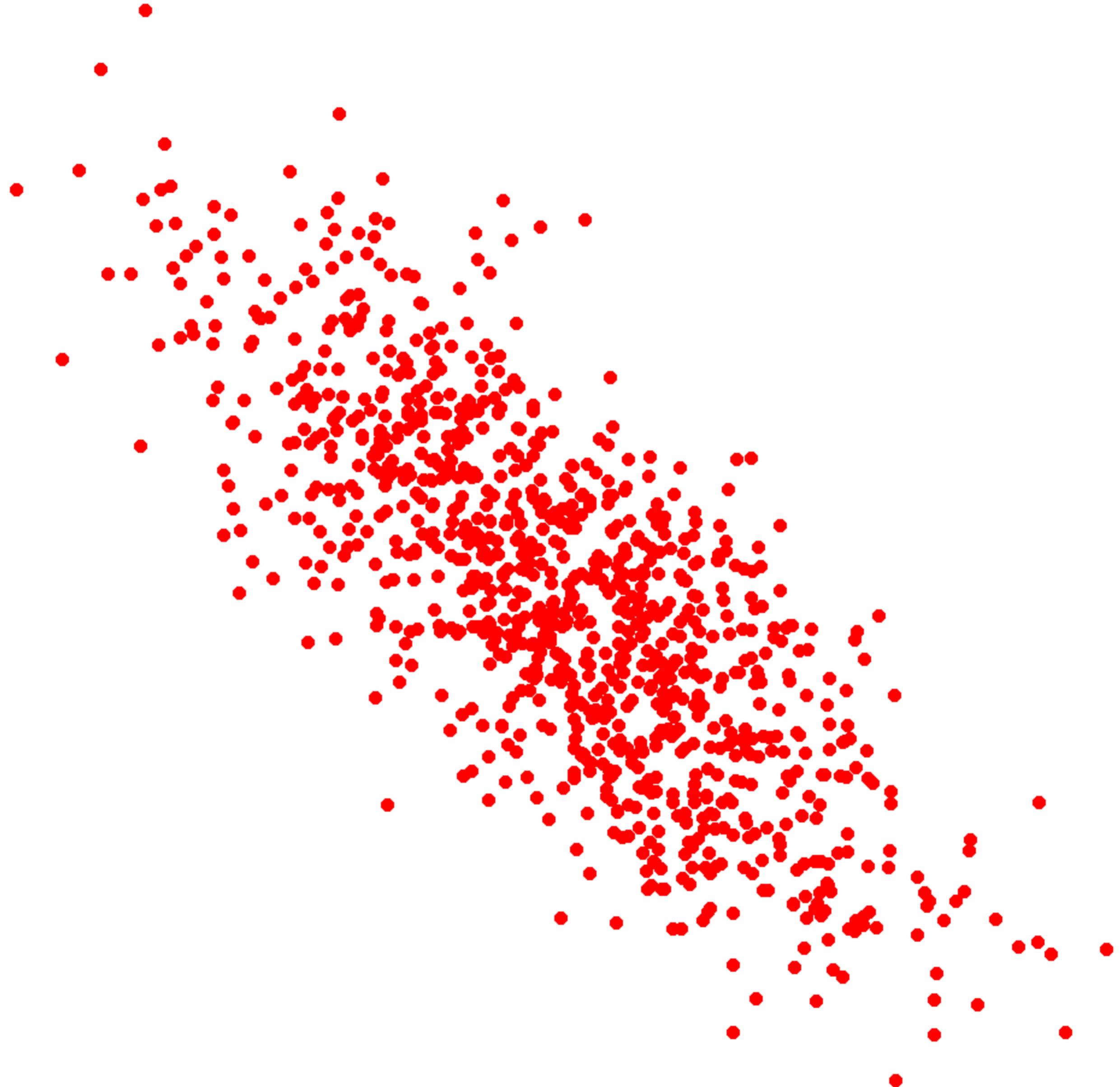
-0.82



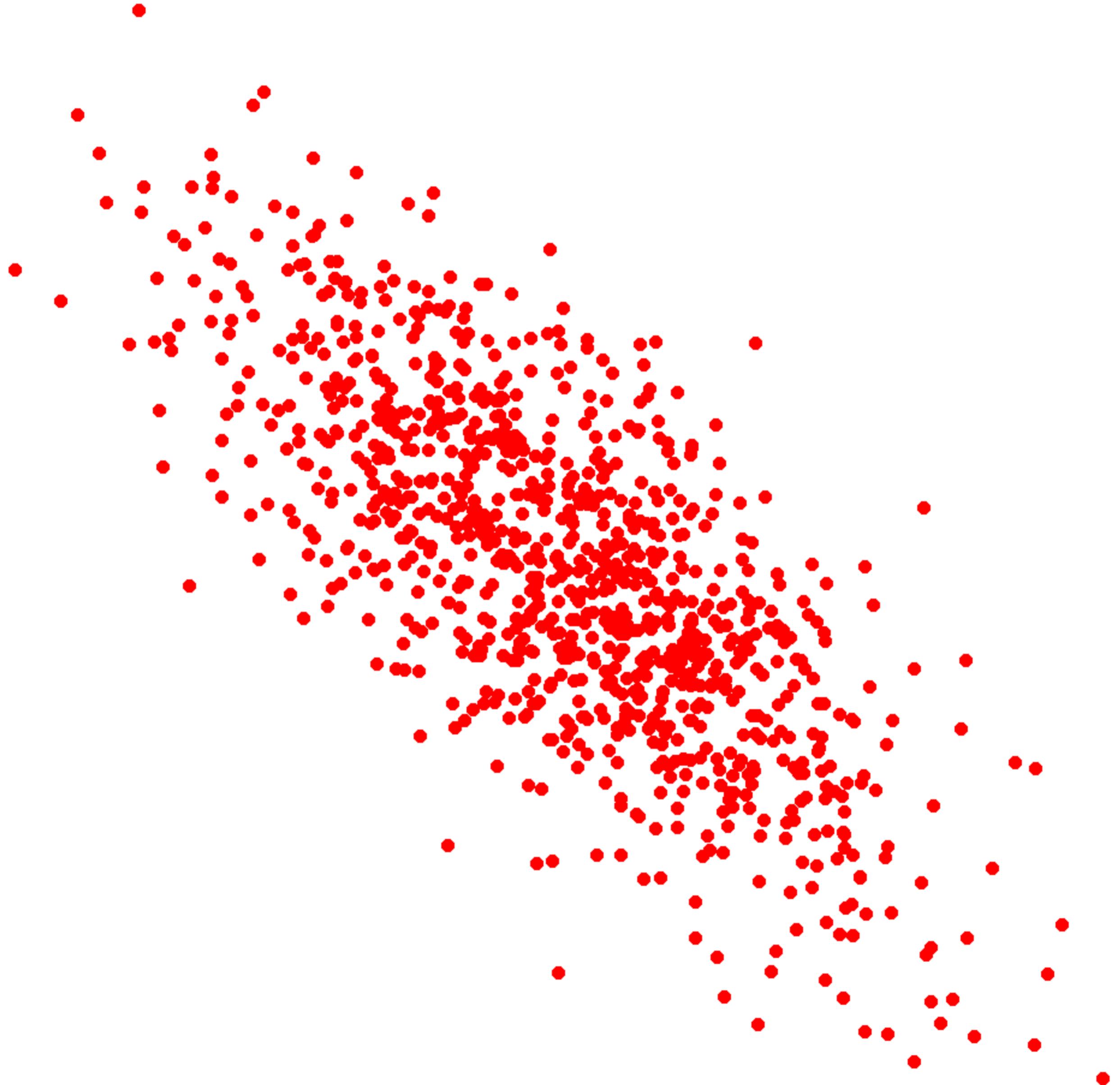
-0.79



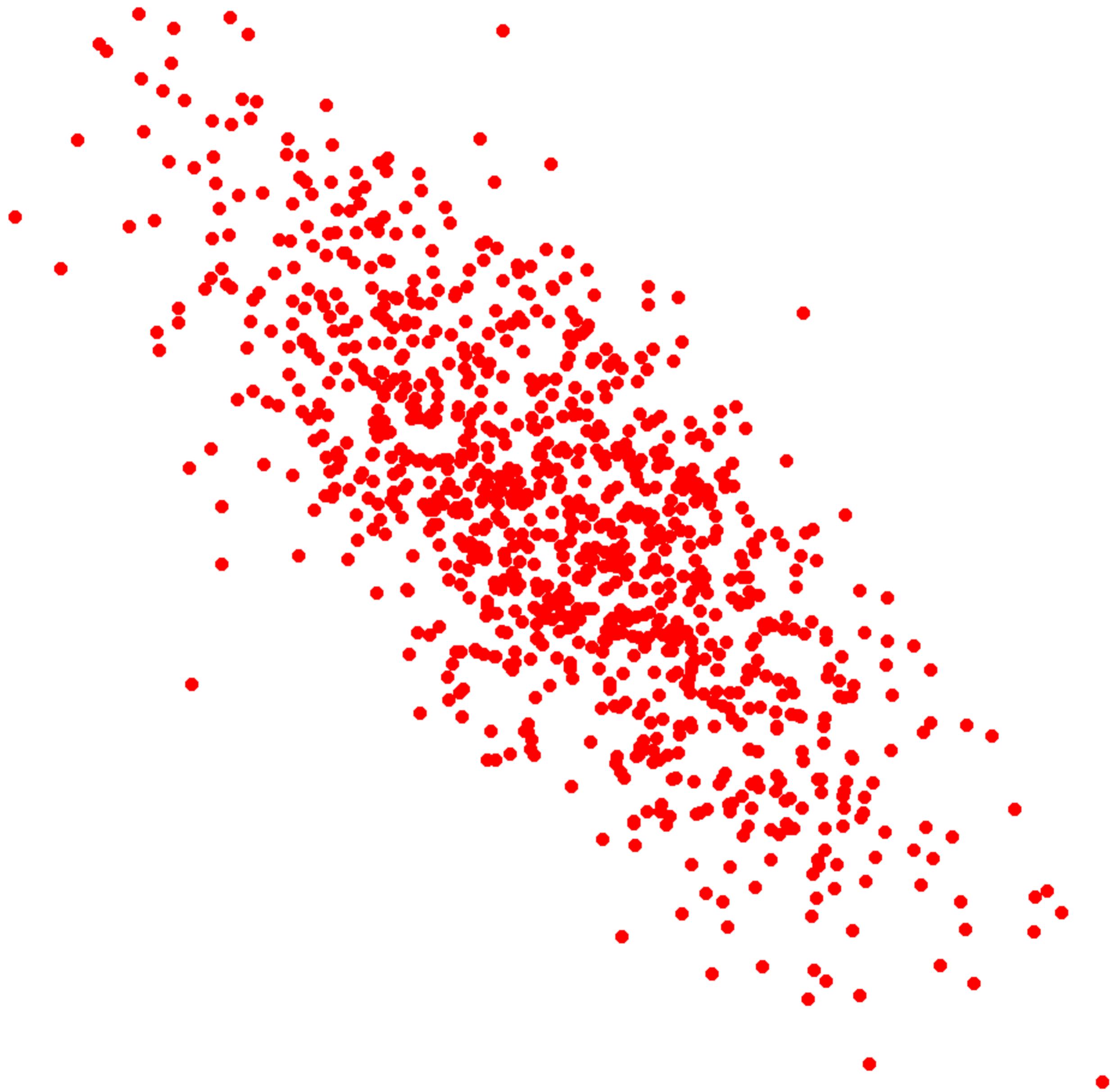
-0.77



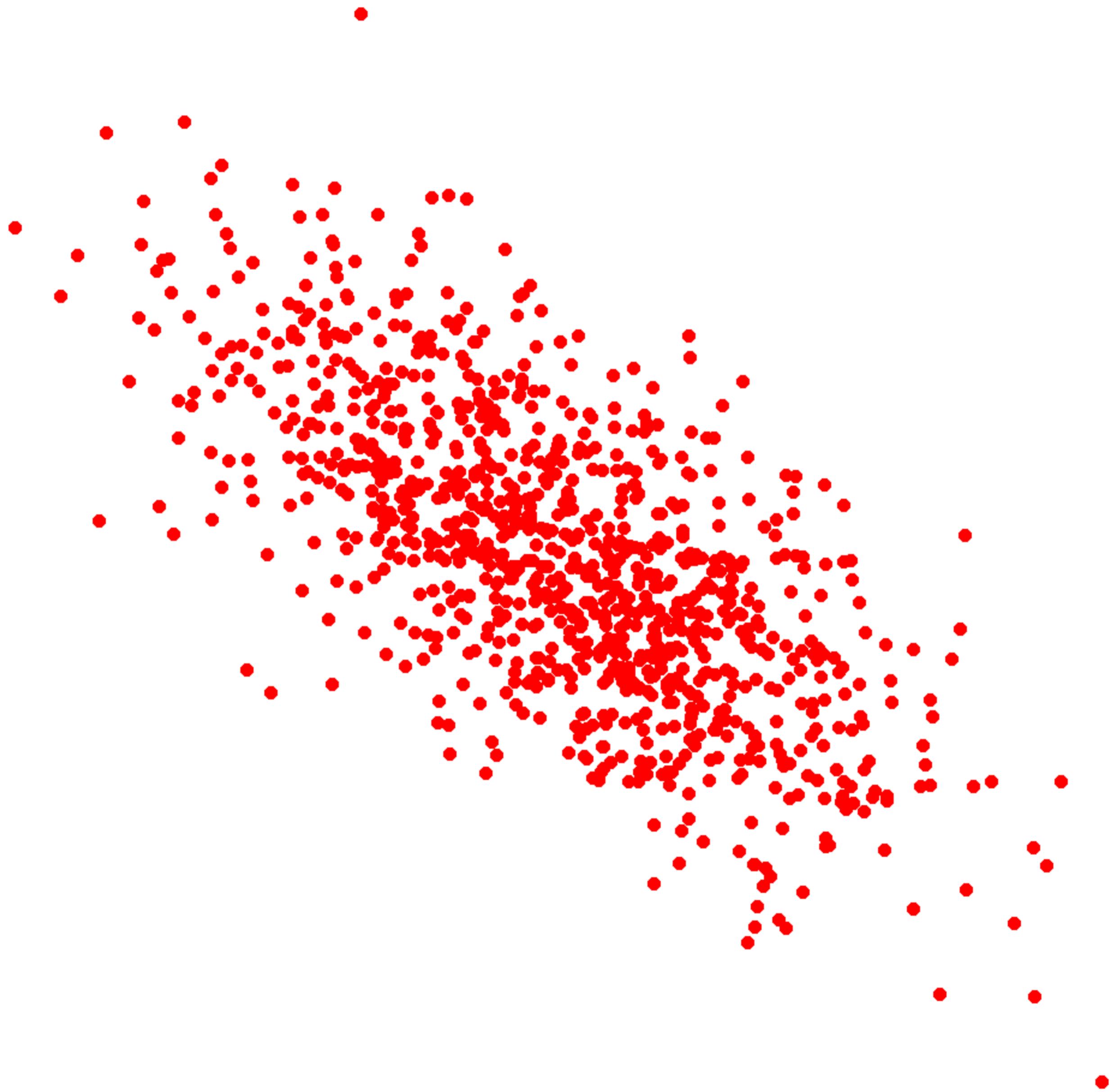
-0.74



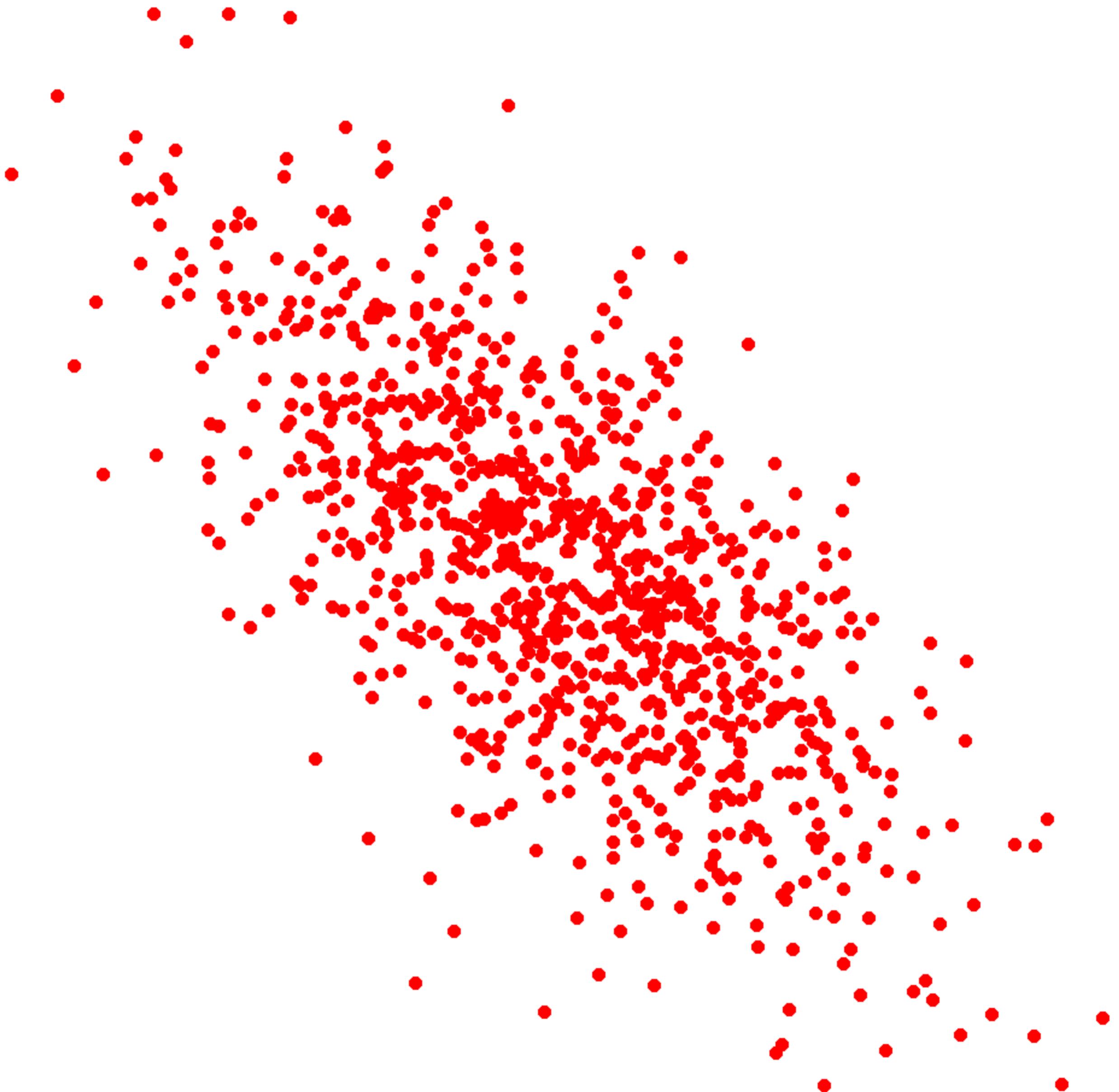
-0.73



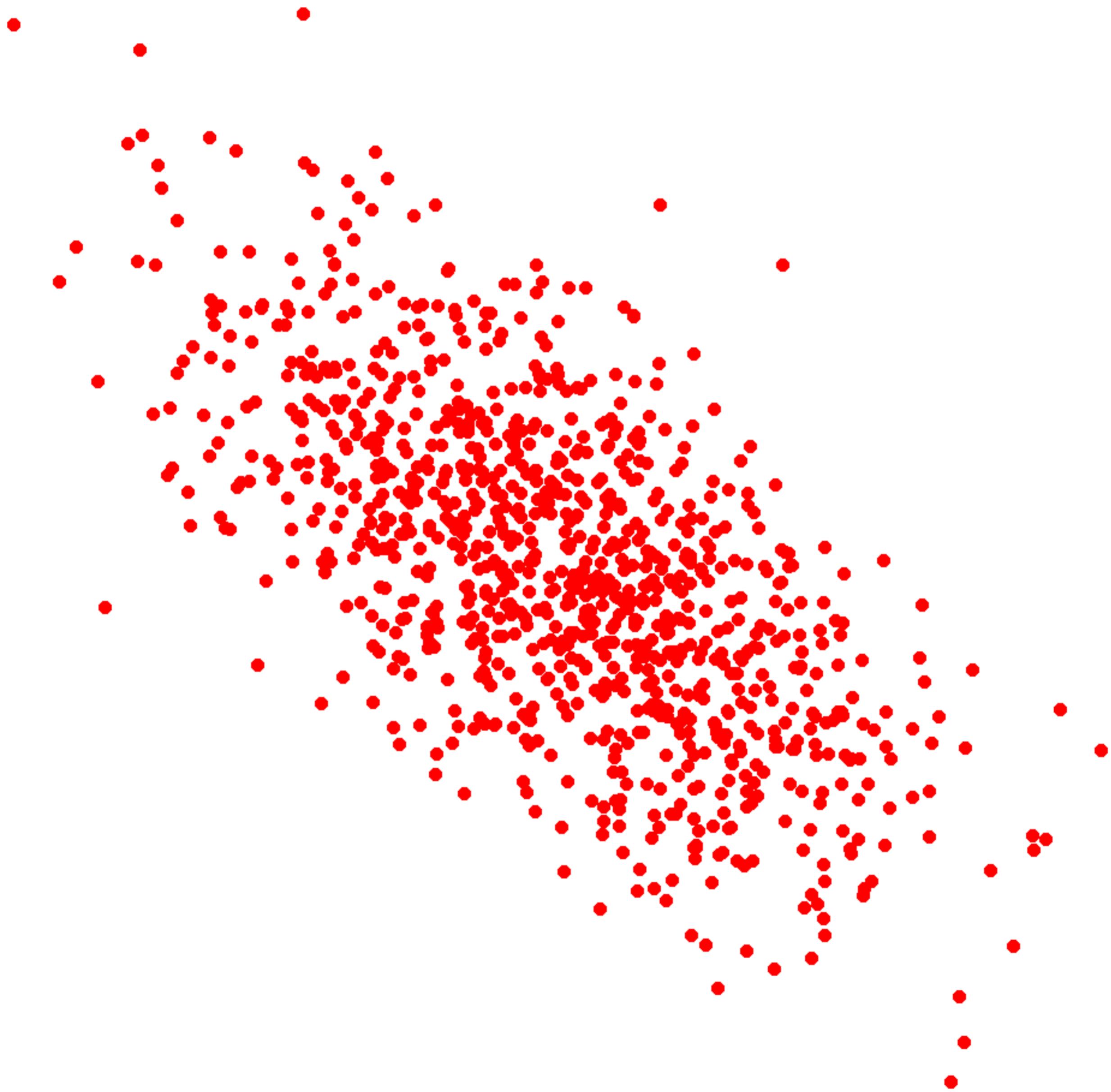
-0.7



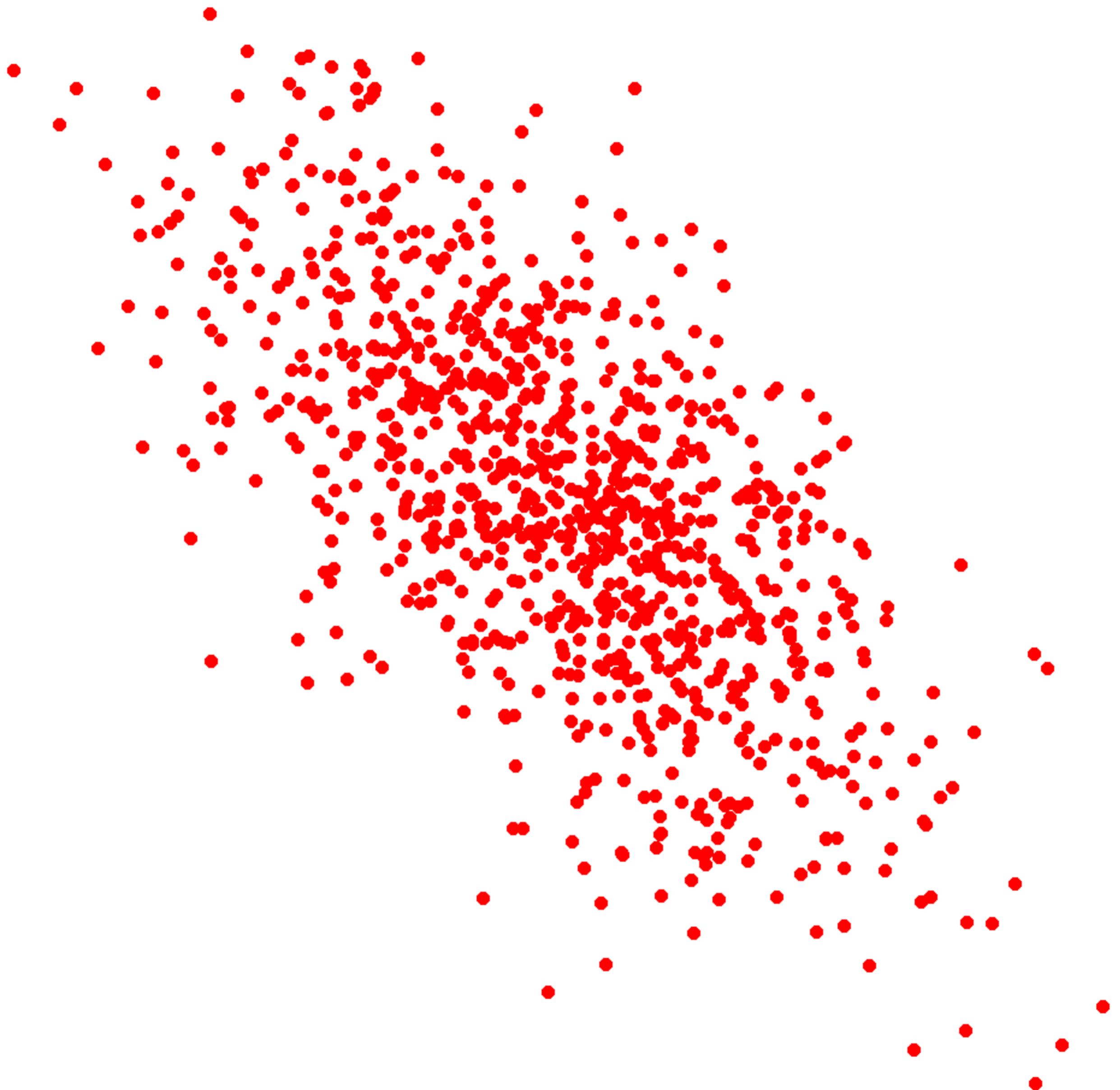
-0.67



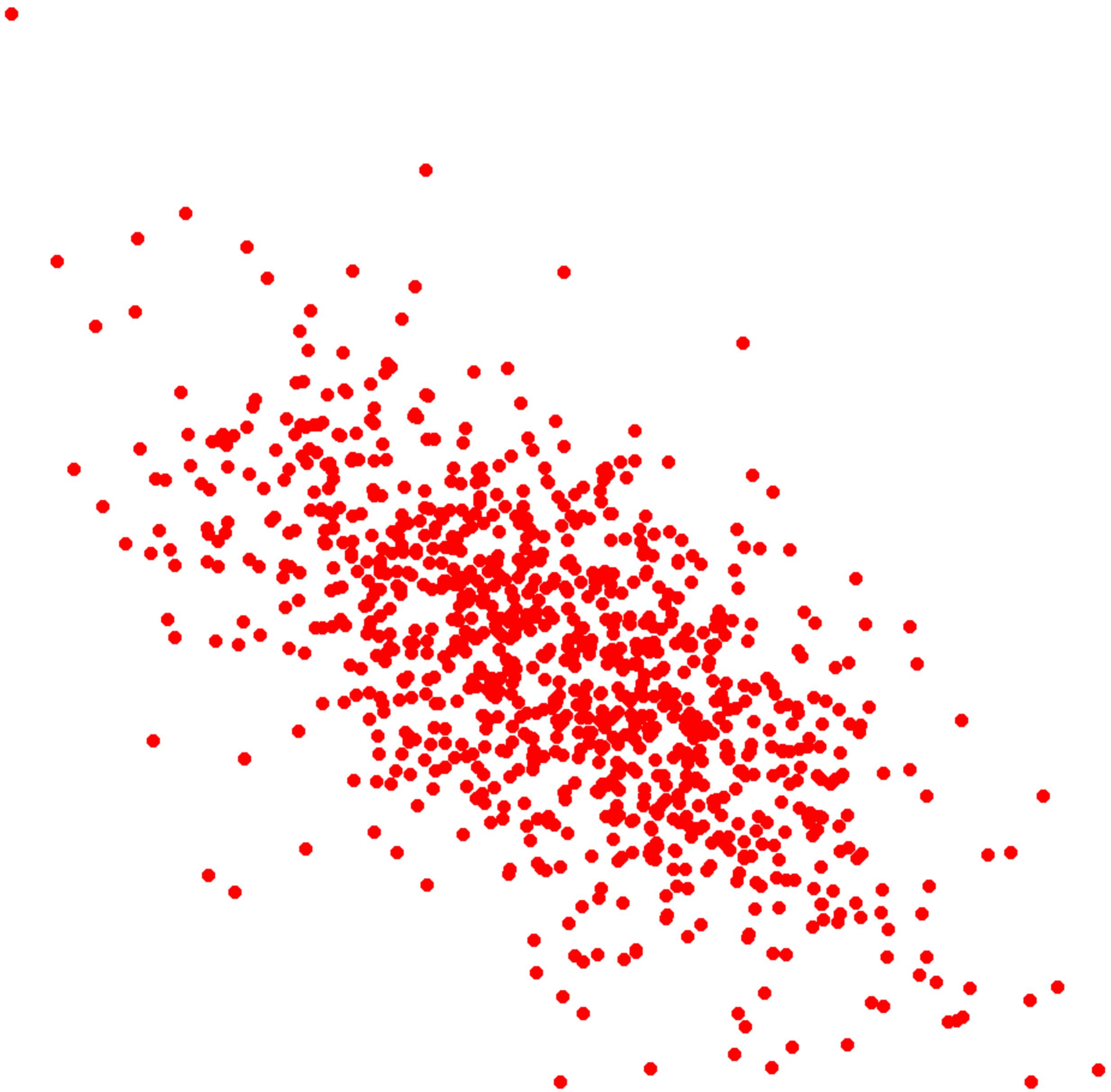
-0.65



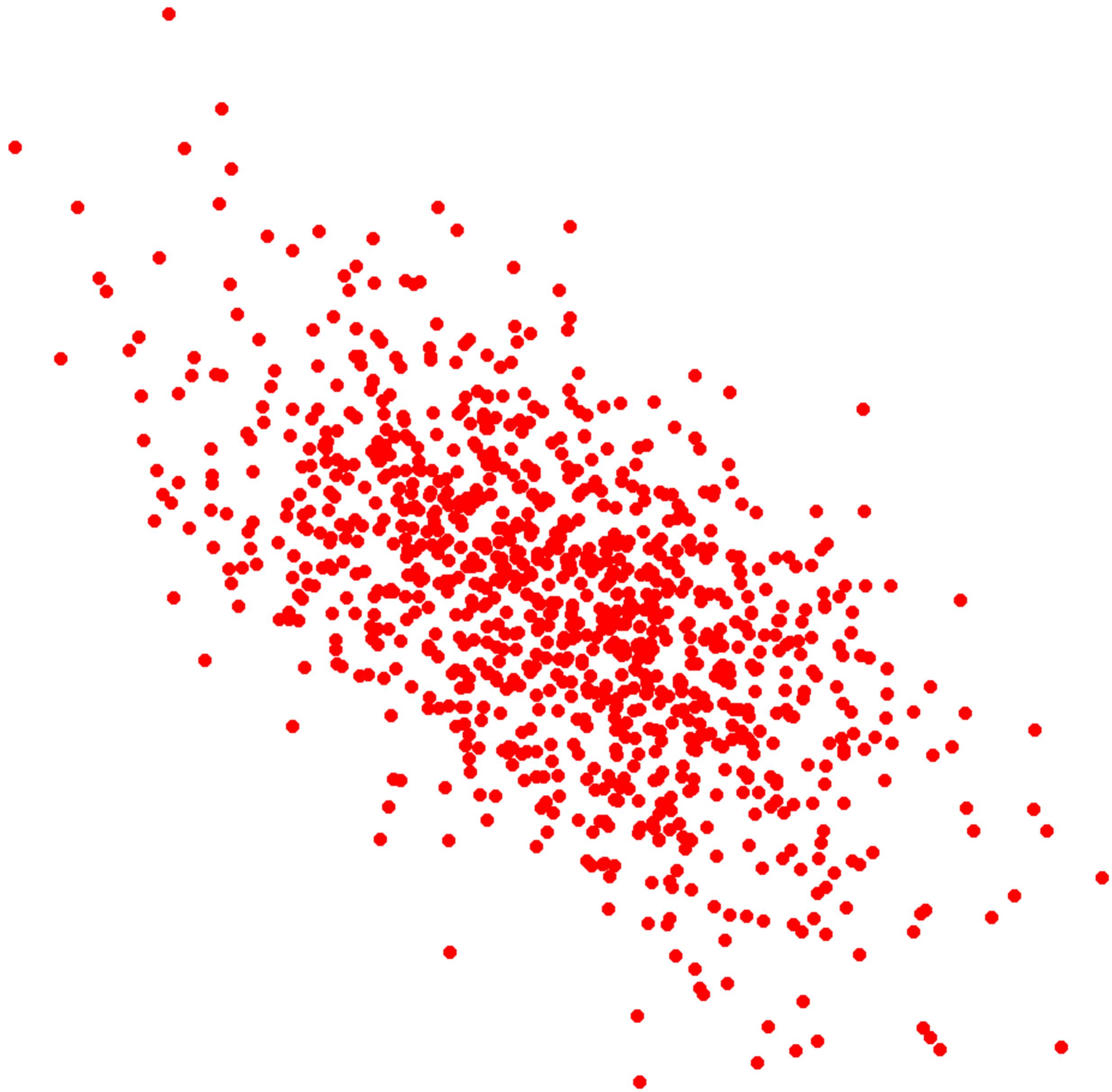
-0.65



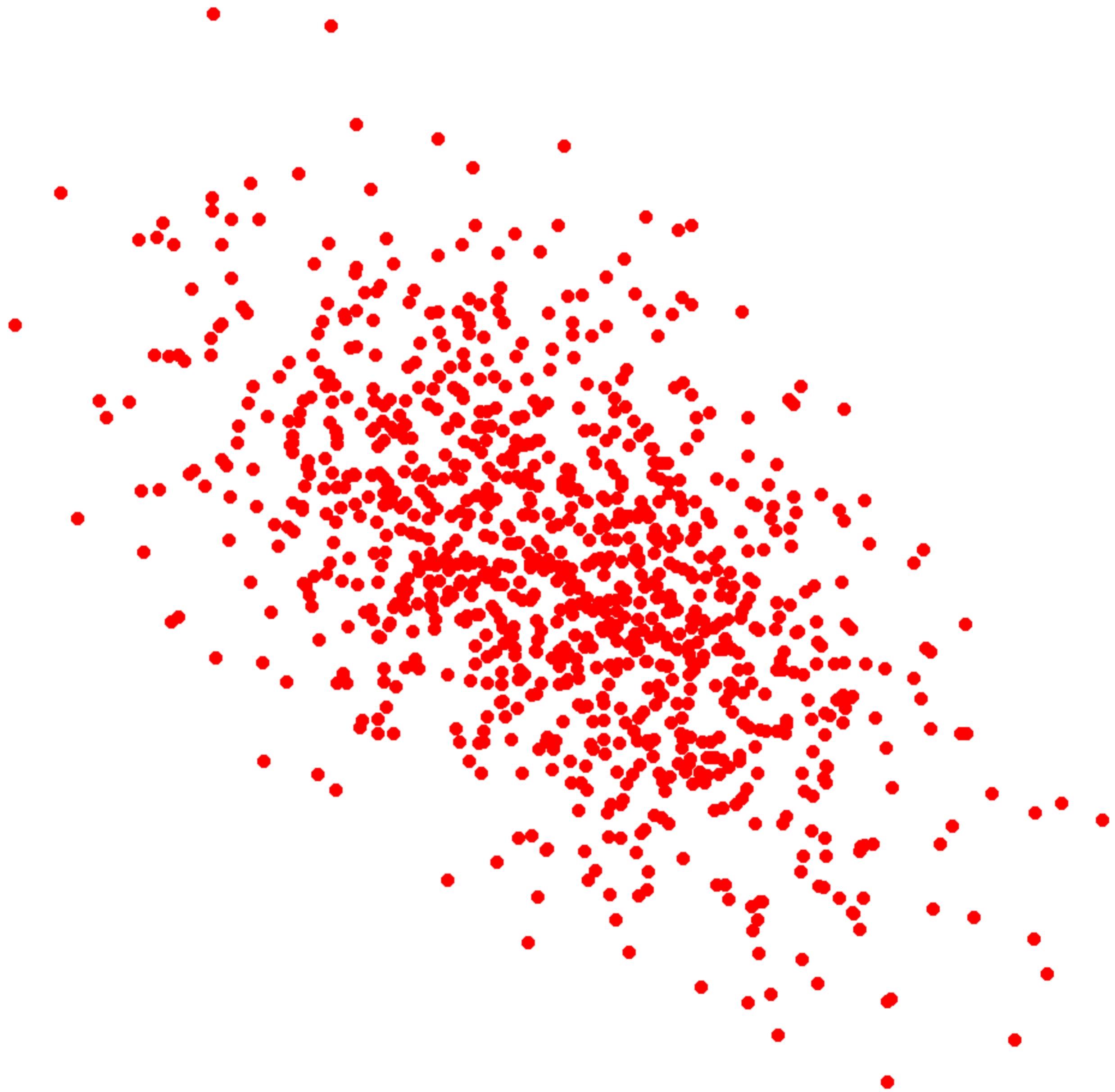
-0.62



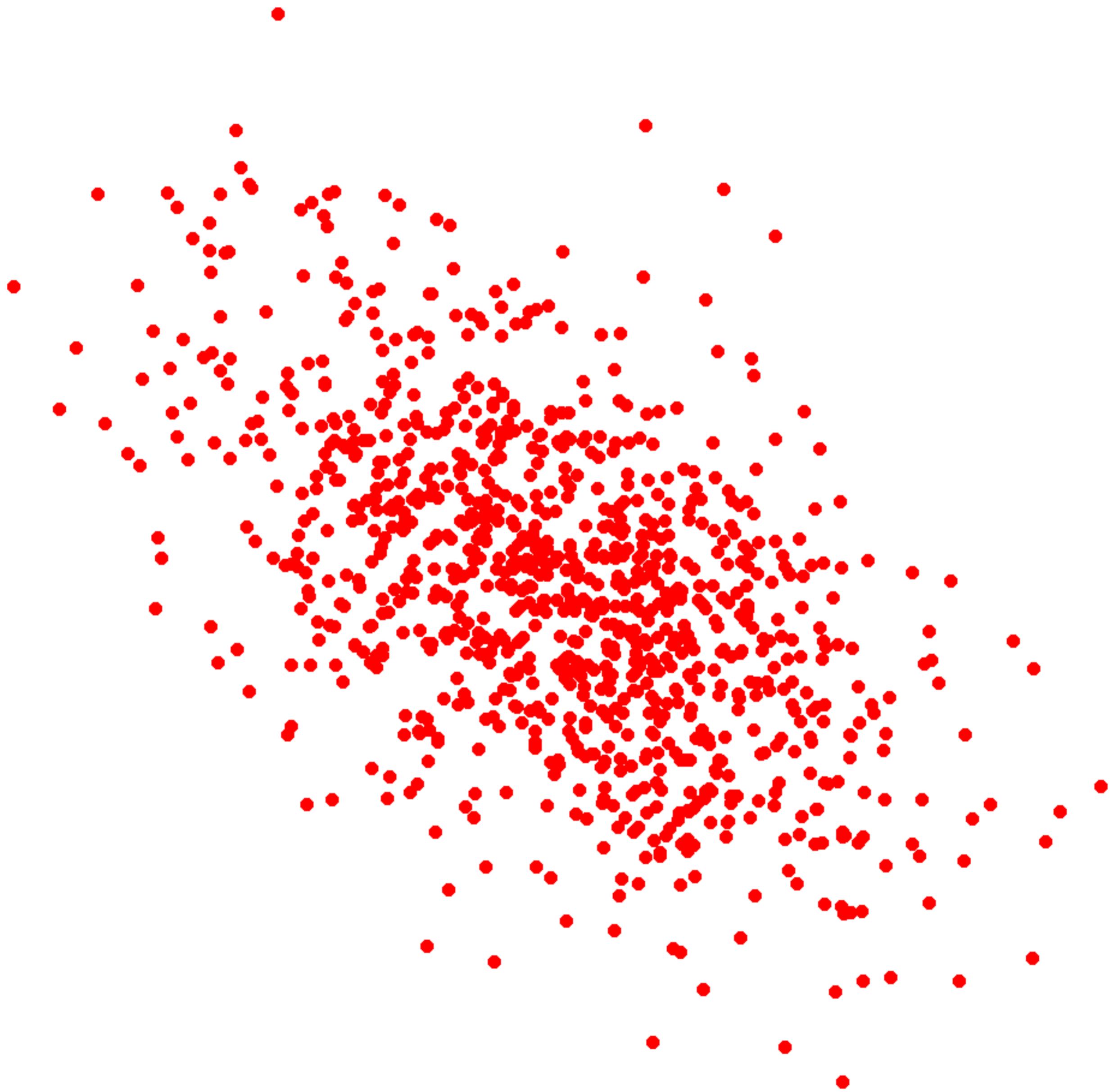
-0.59



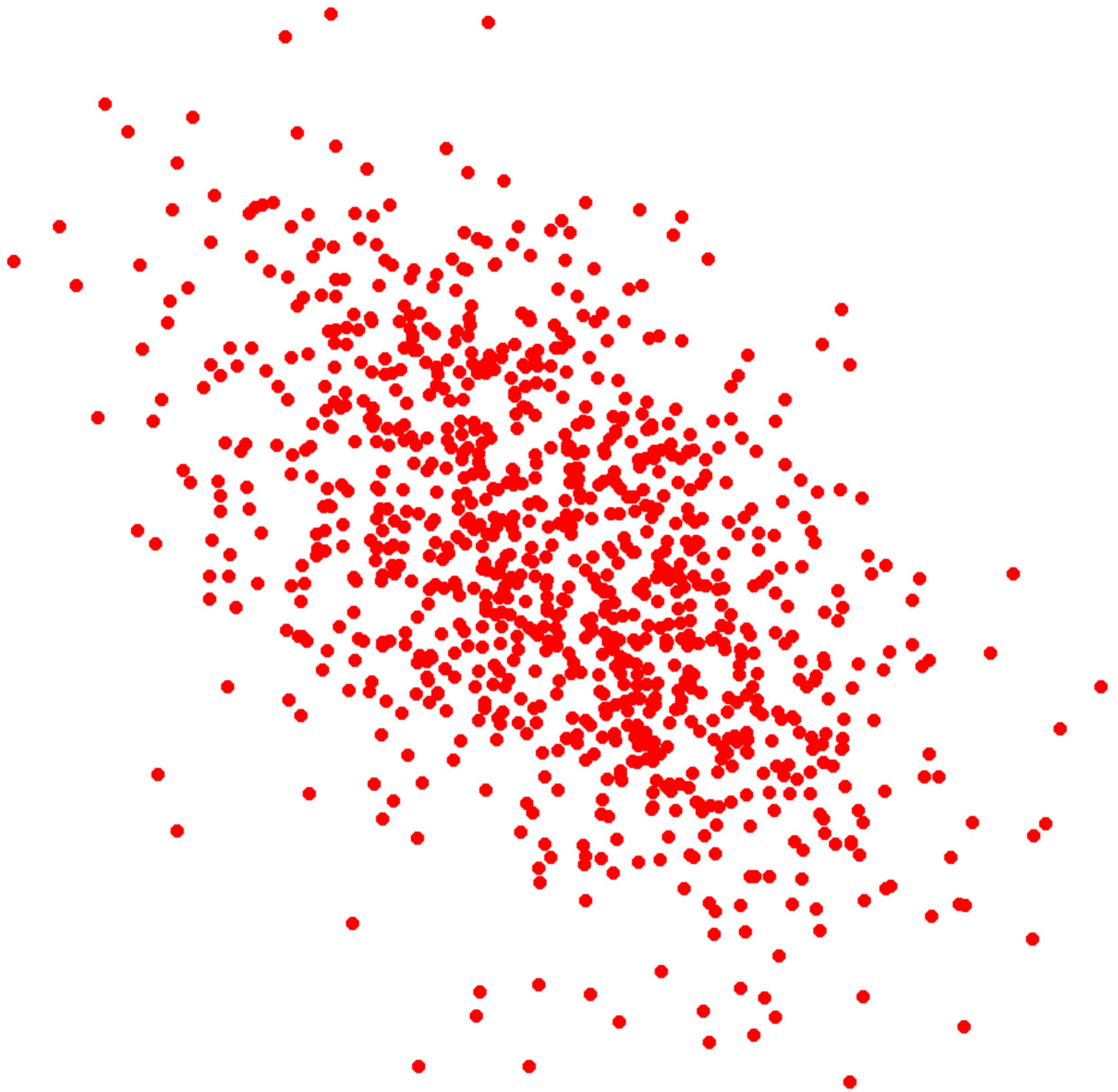
-0.53



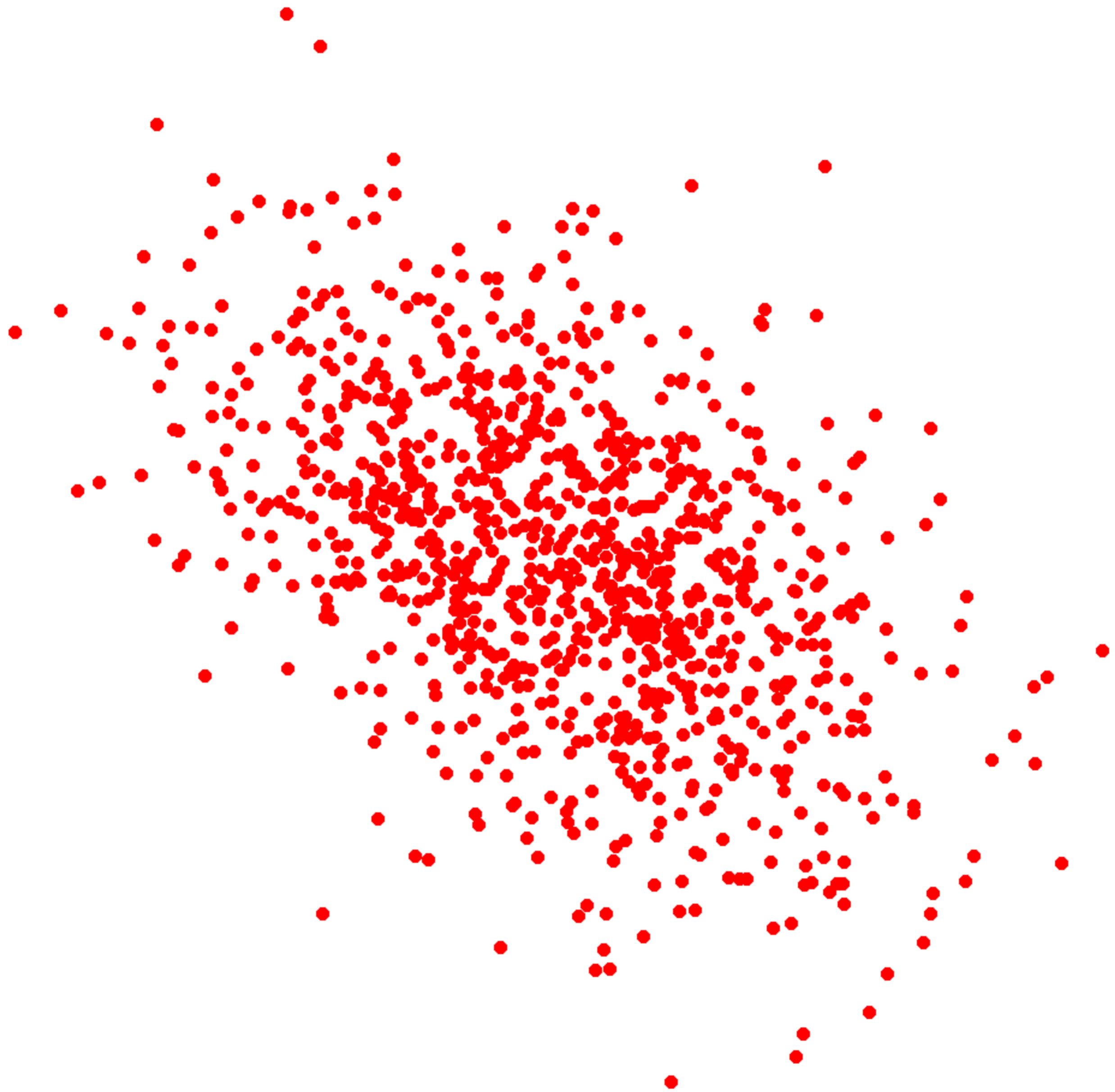
-0.54



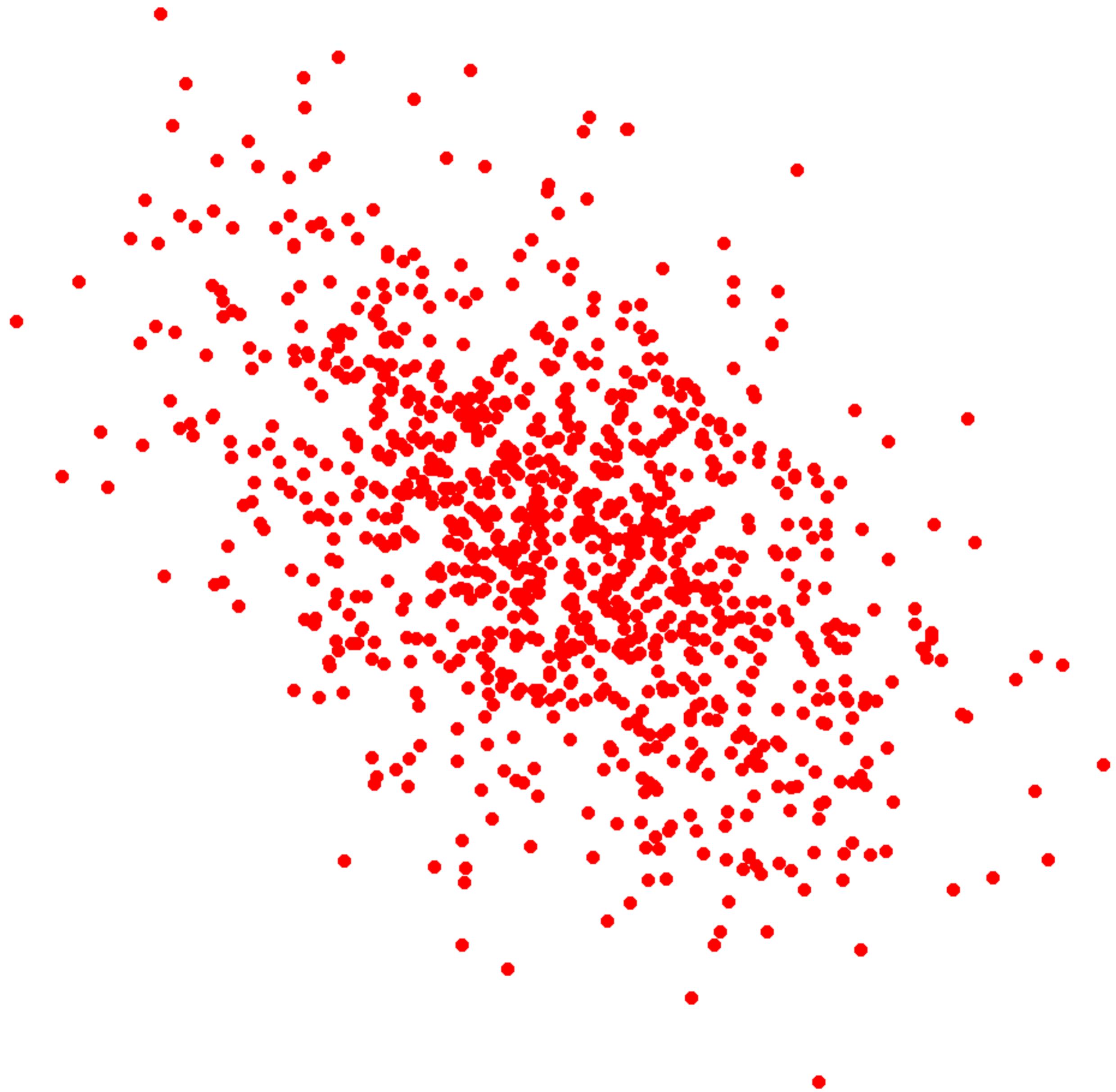
-0.49



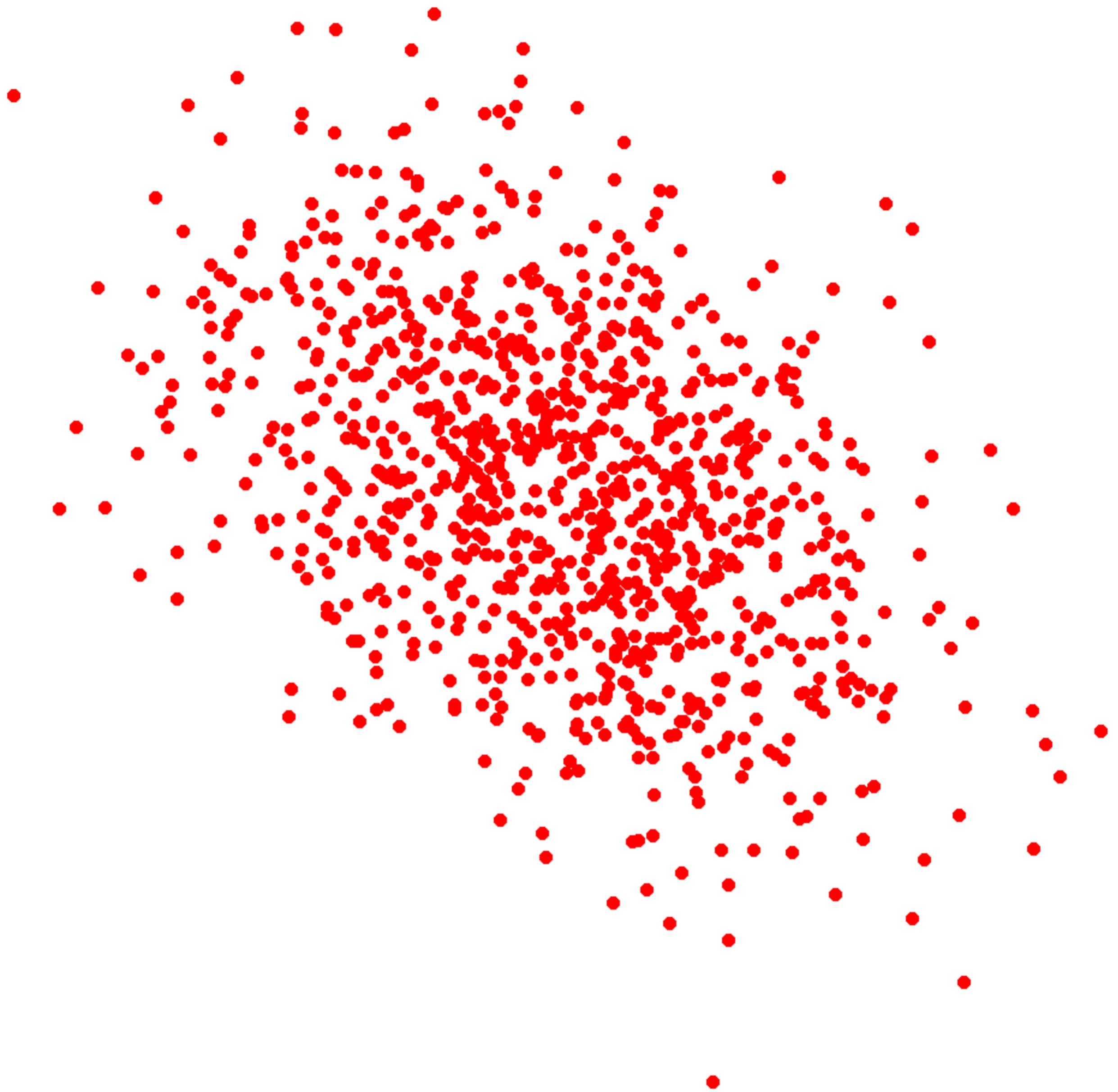
-0.47



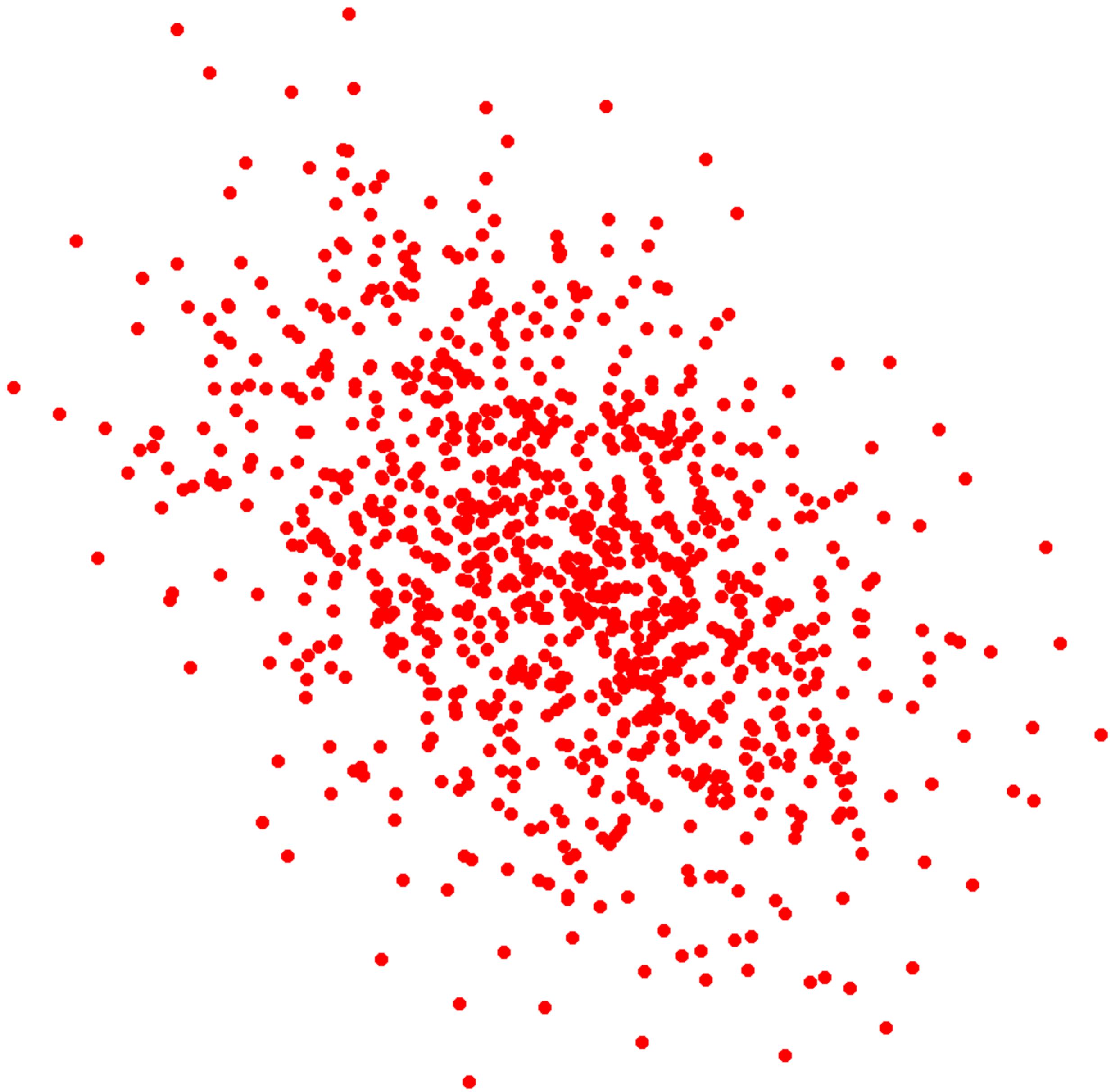
-0.47



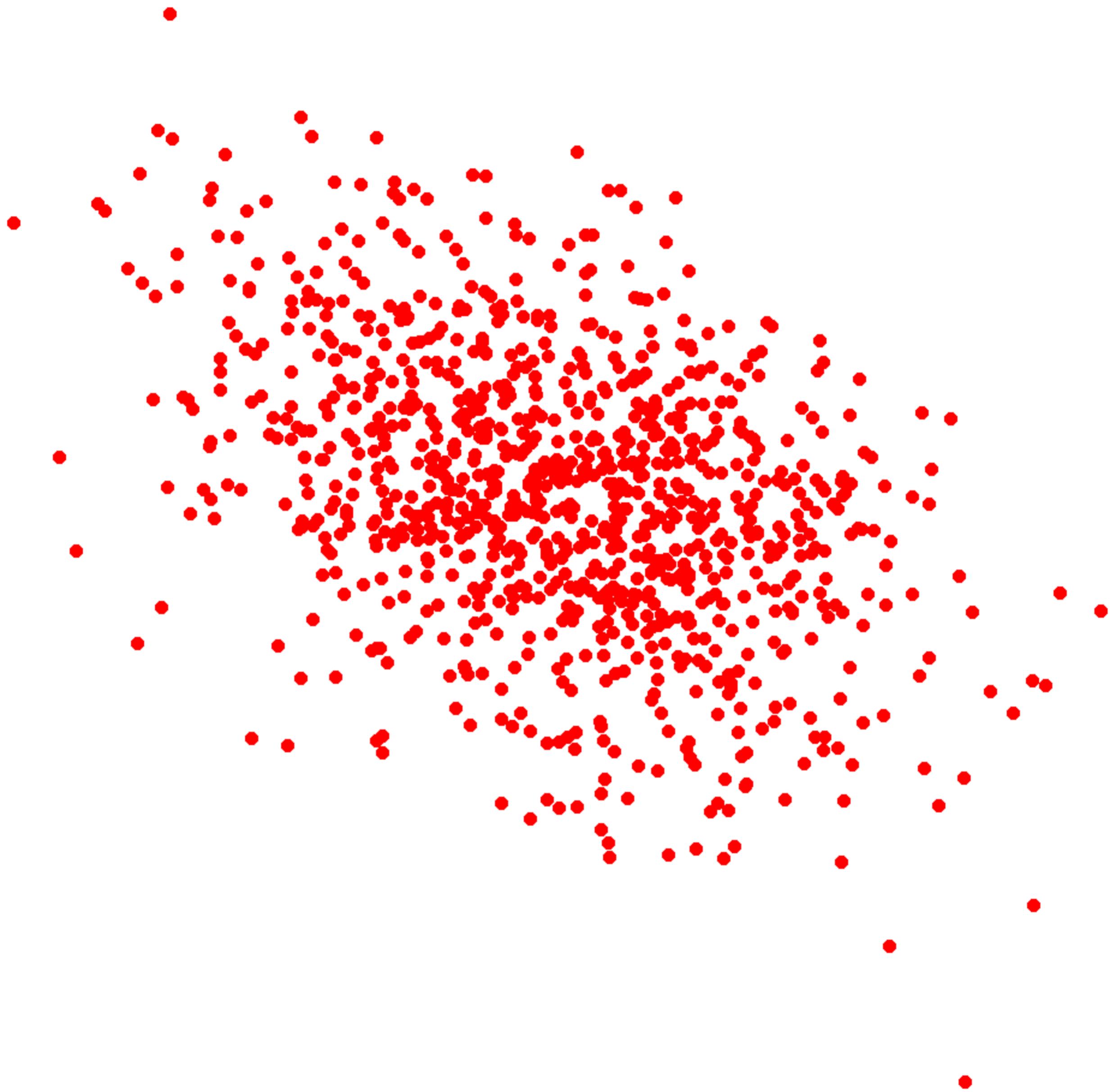
-0.4



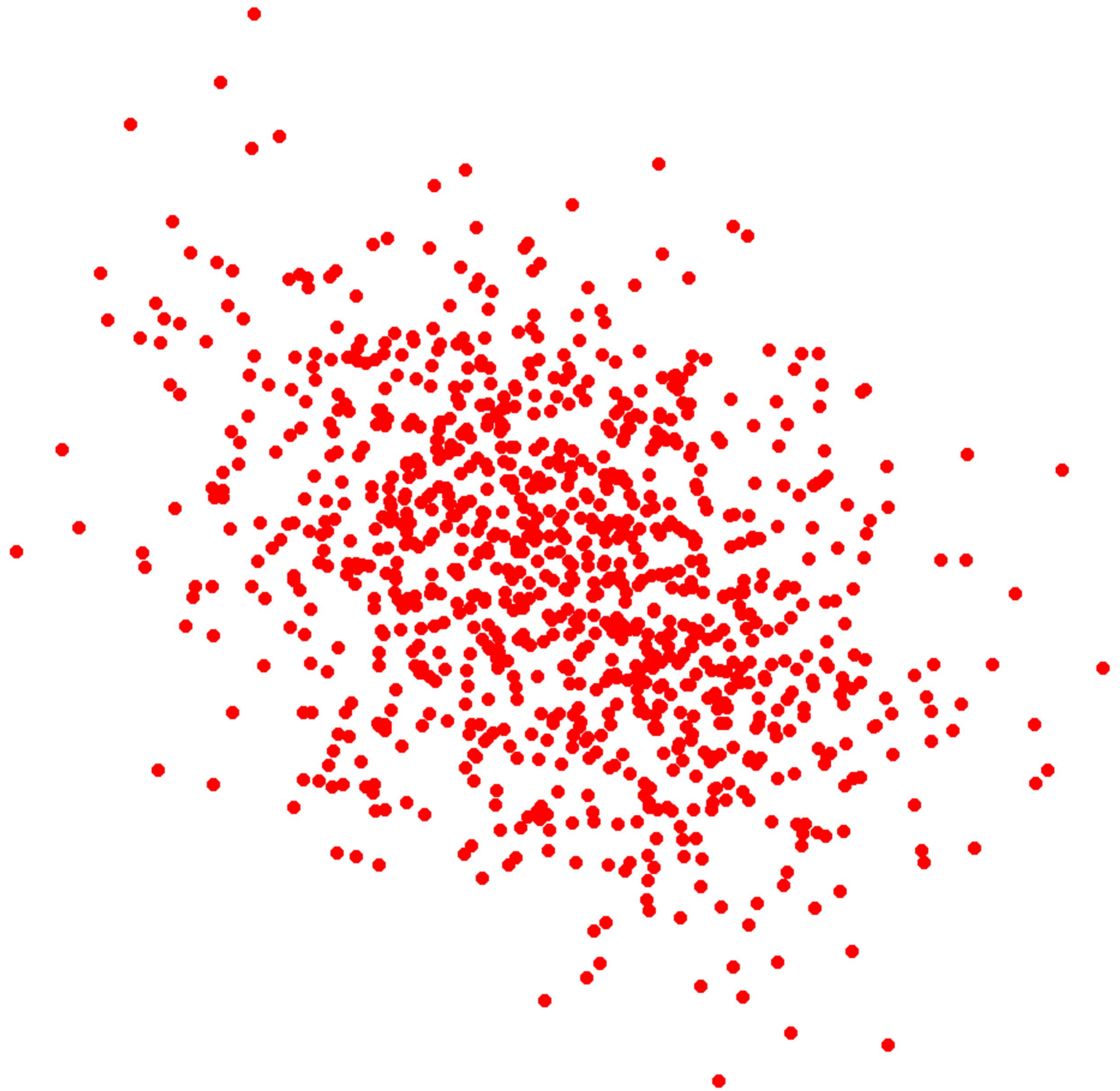
-0.41



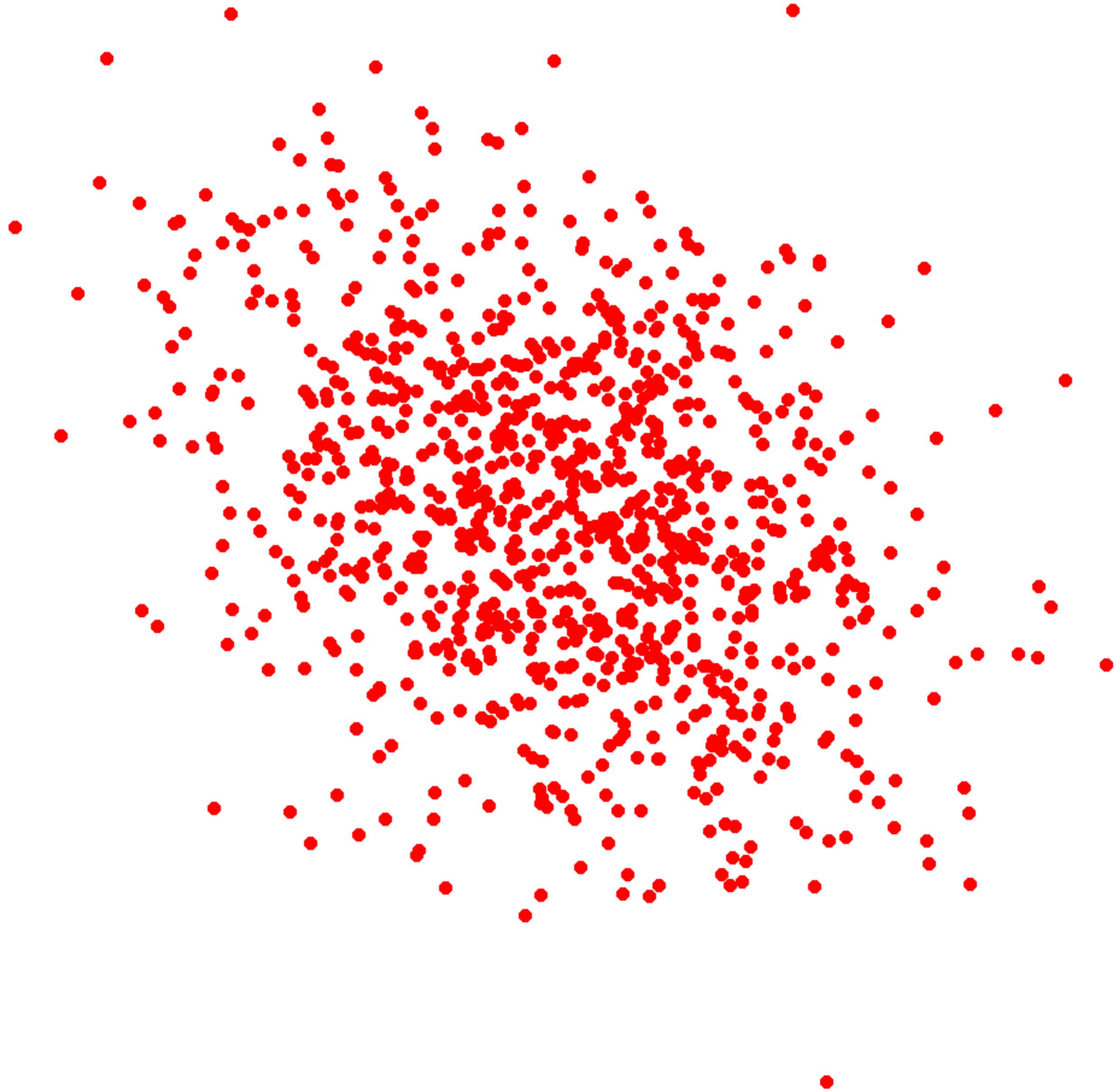
-0.41



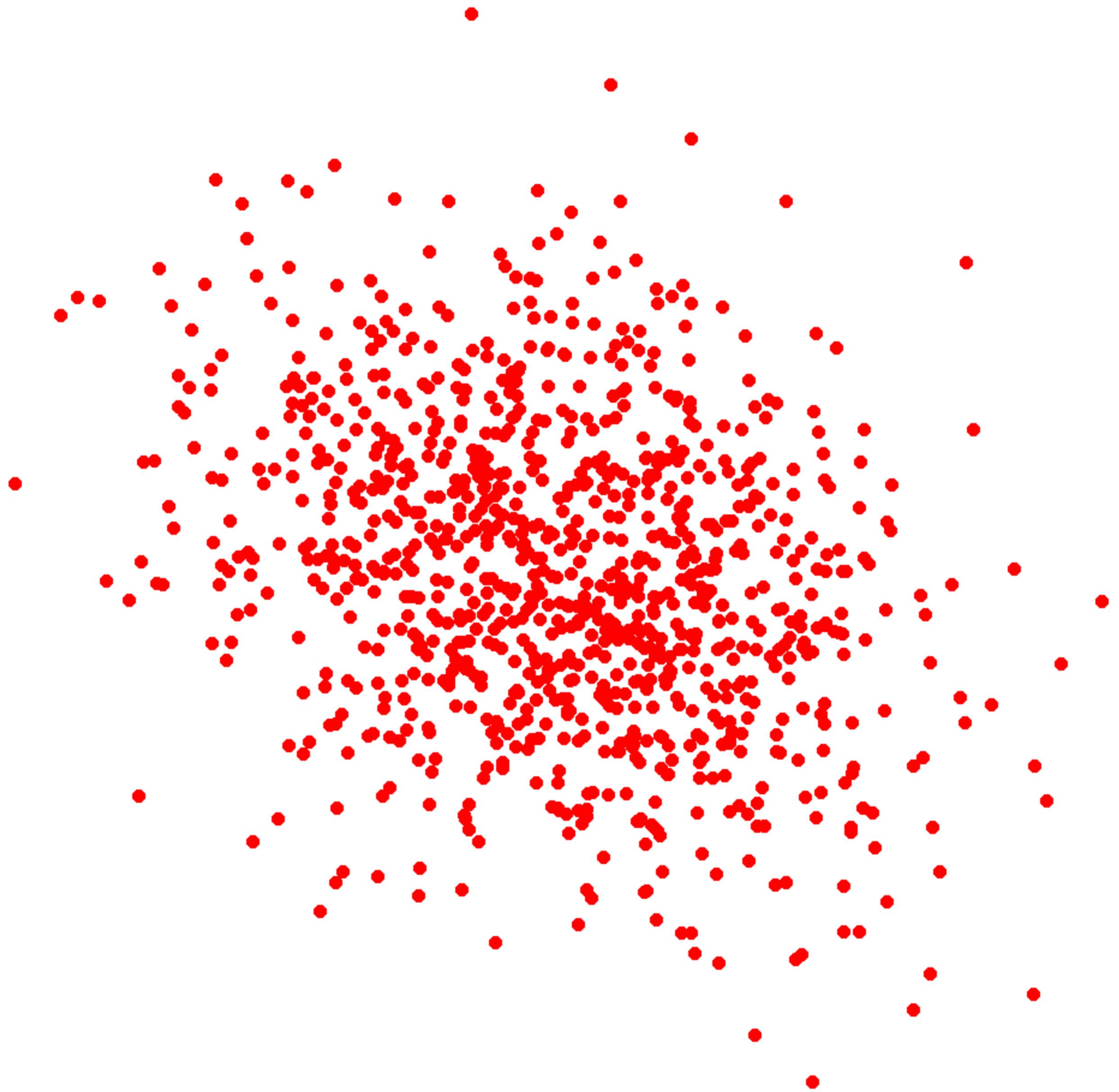
-0.35



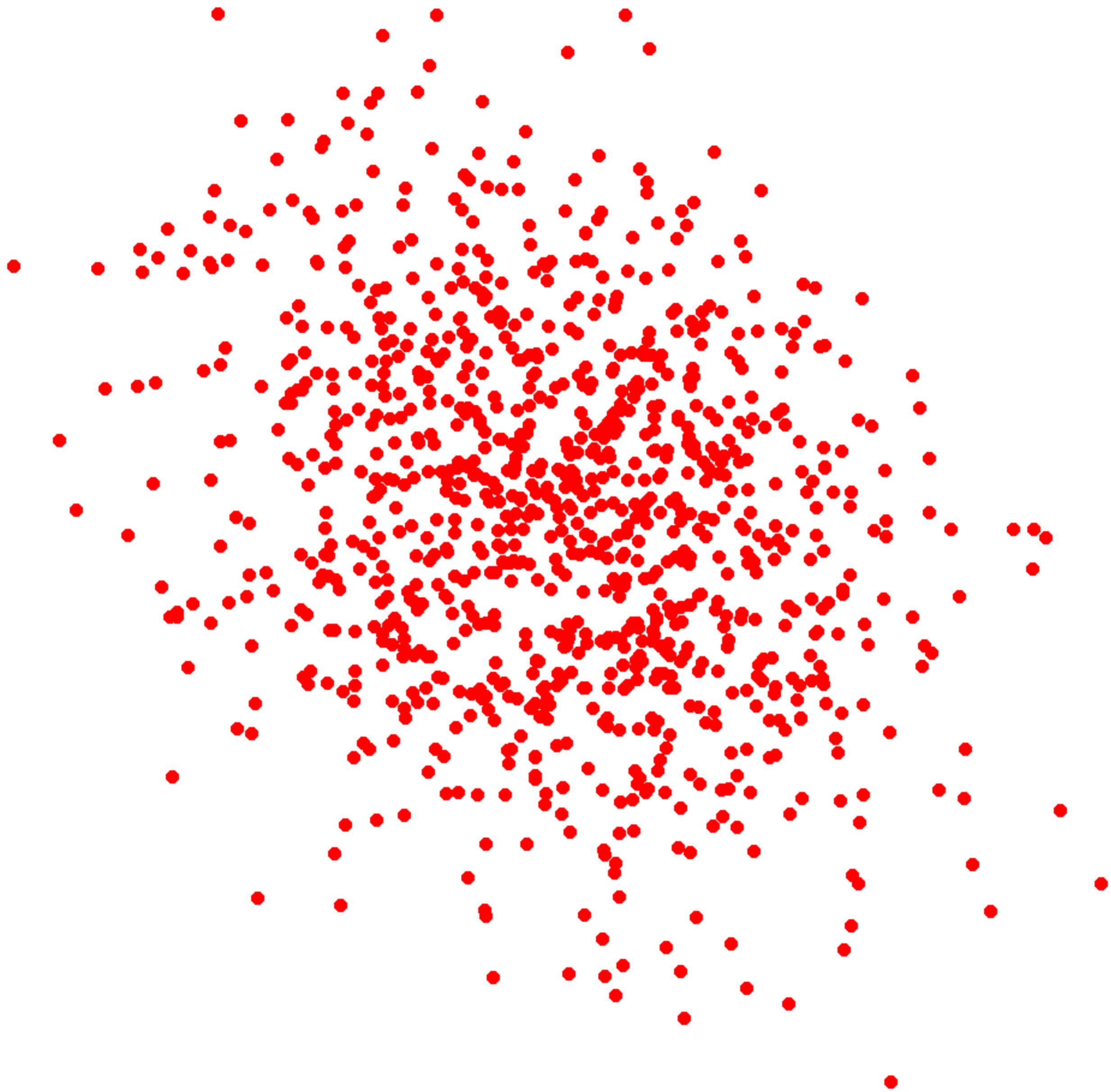
-0.32



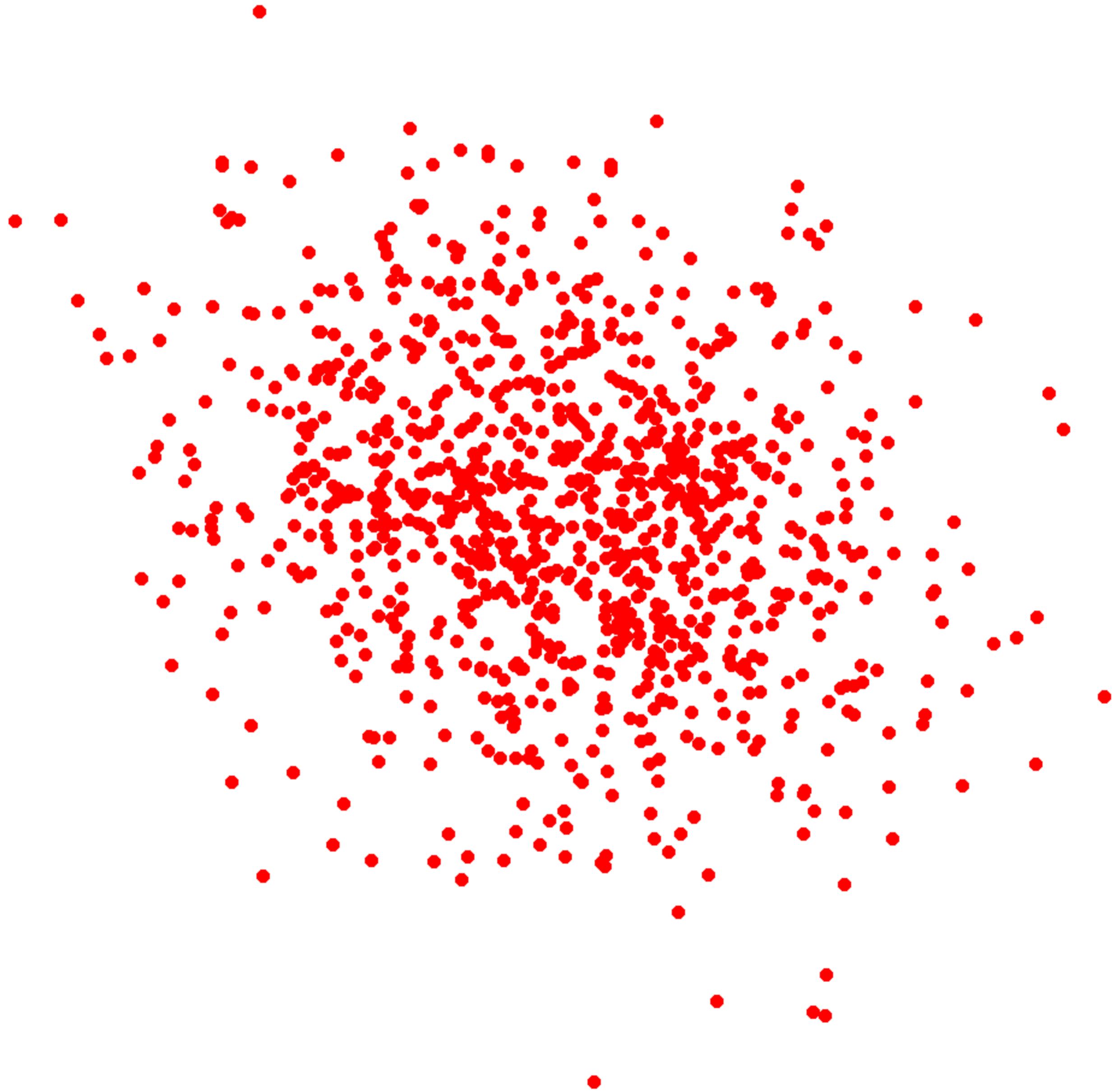
-0.3



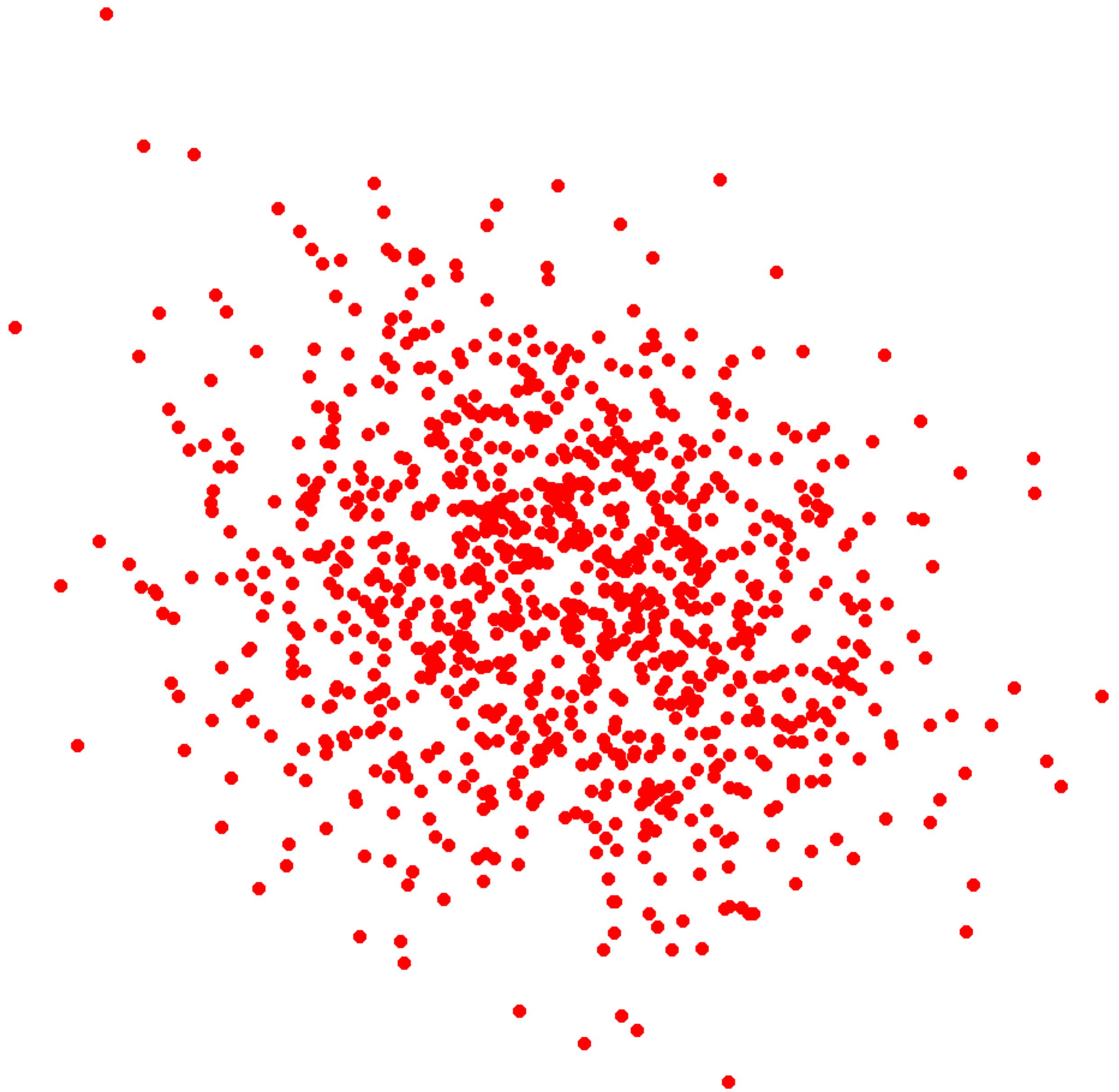
-0.26



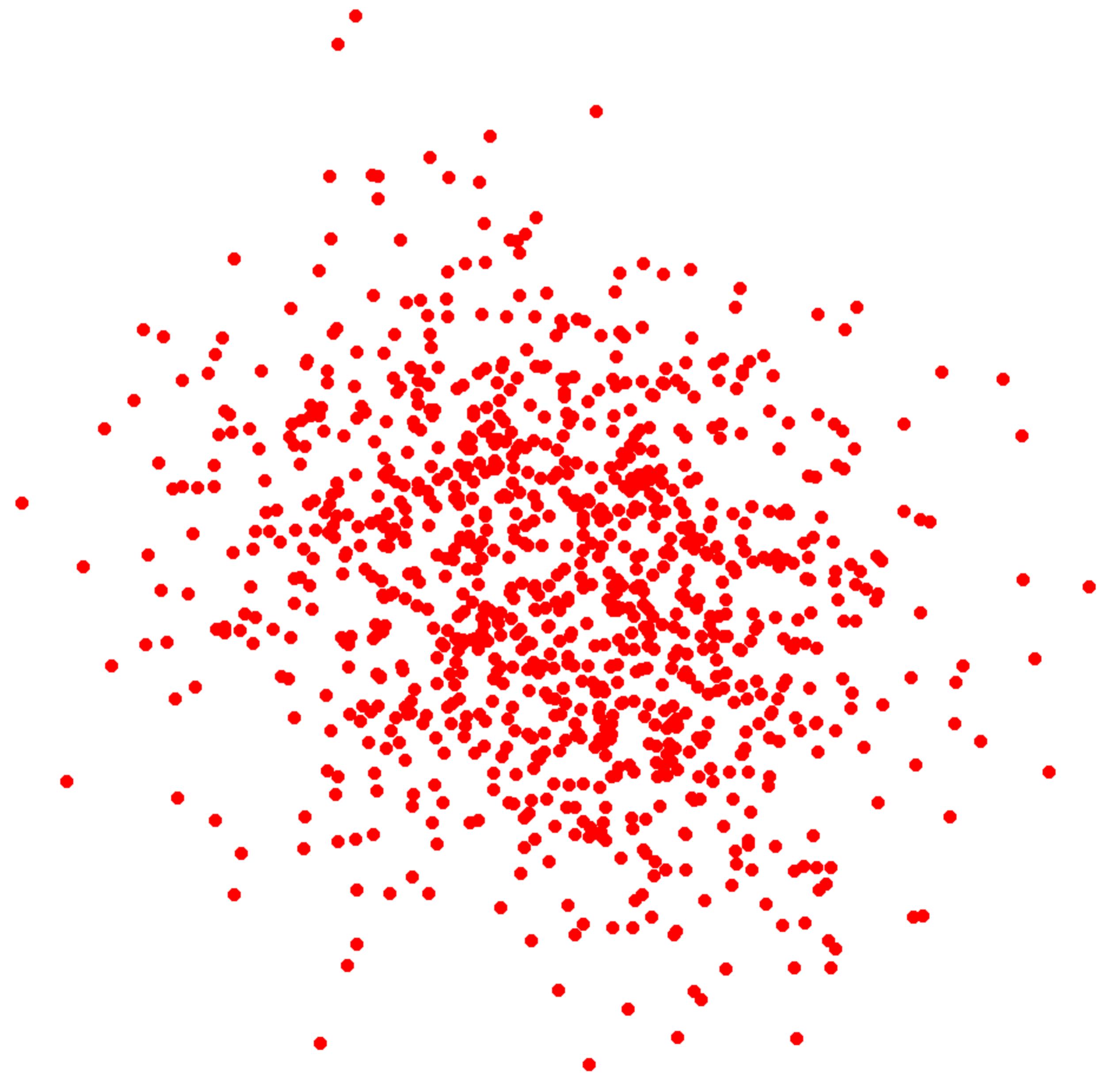
-0.24



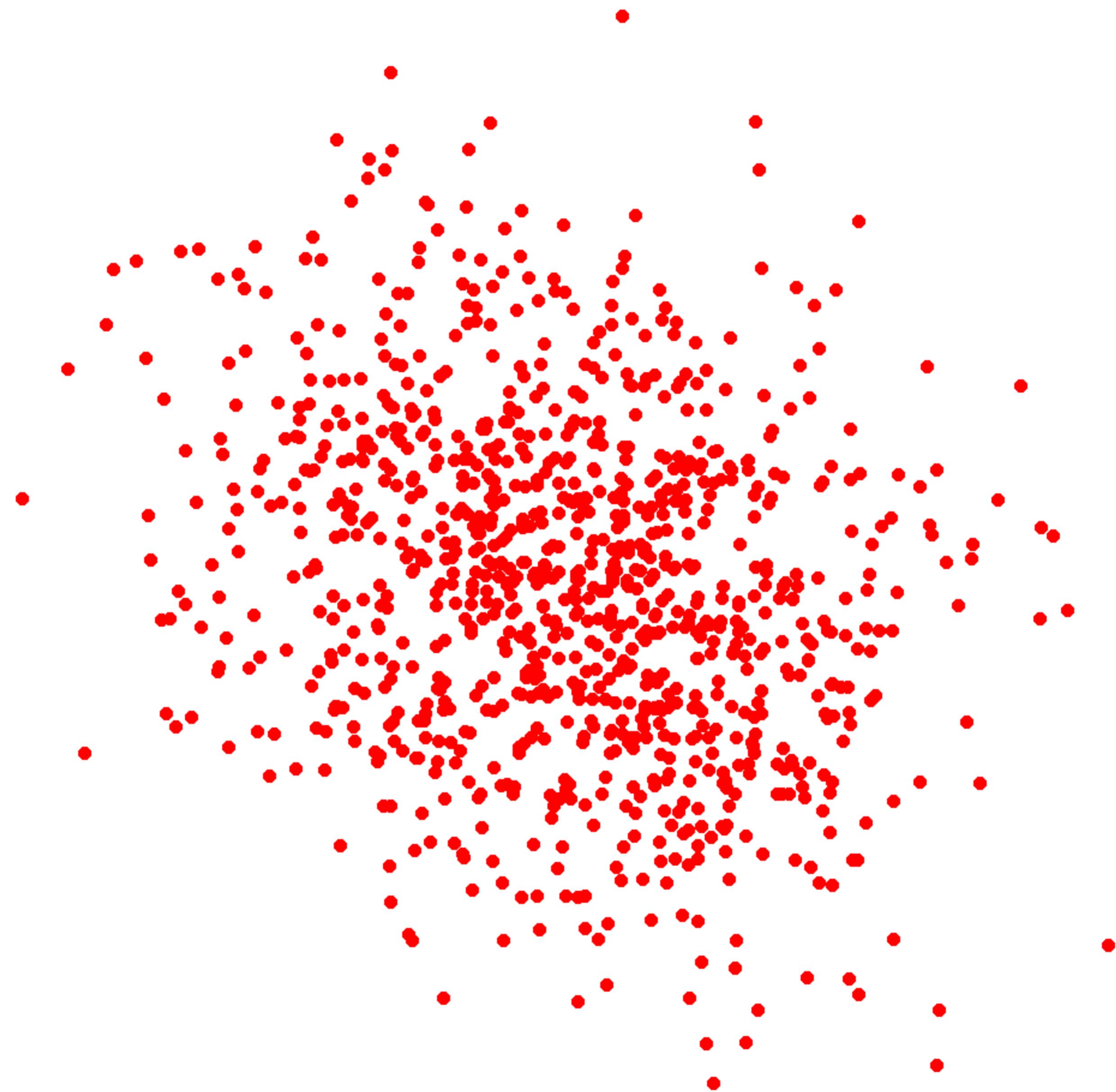
-0.2



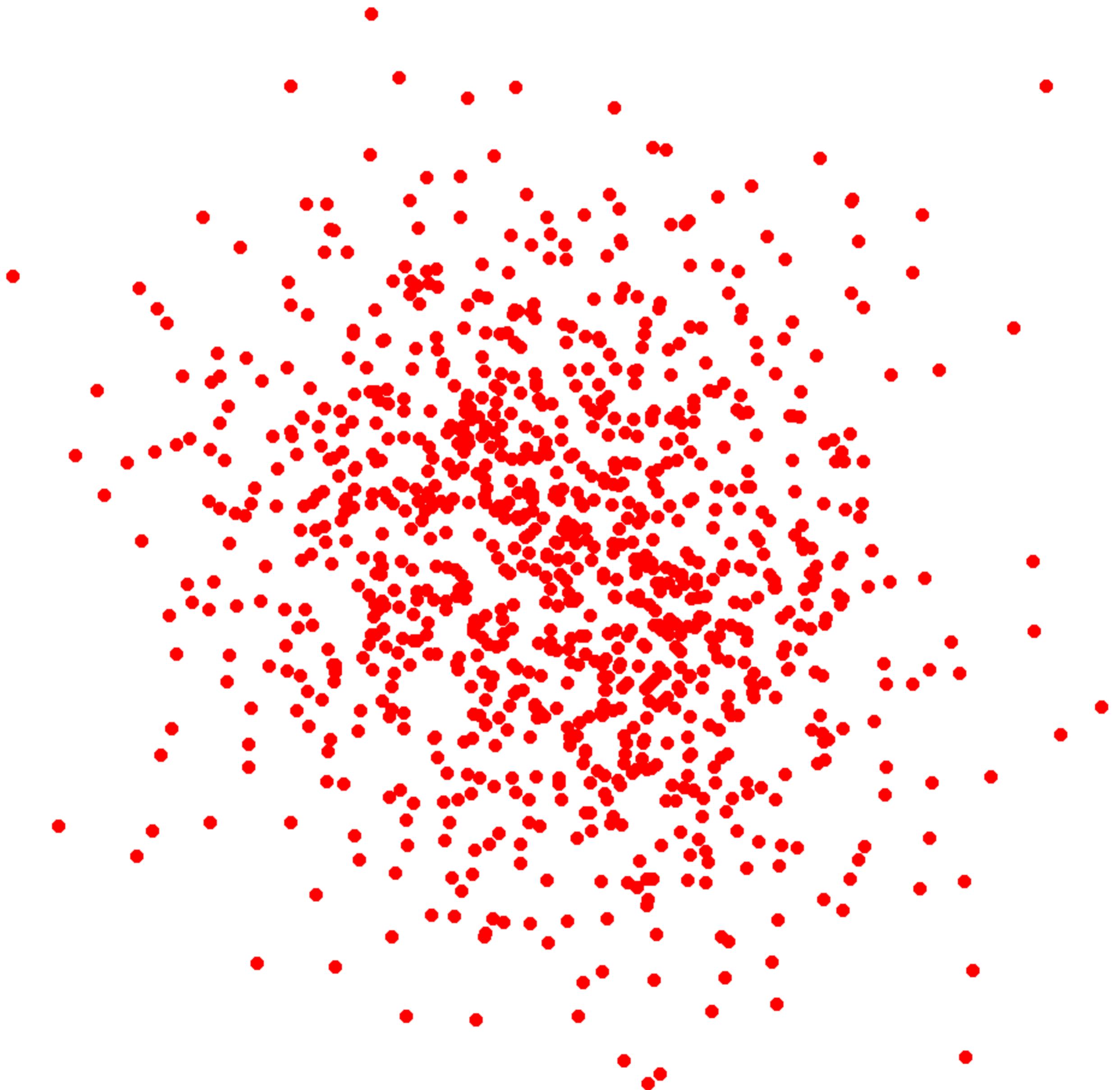
-0.21



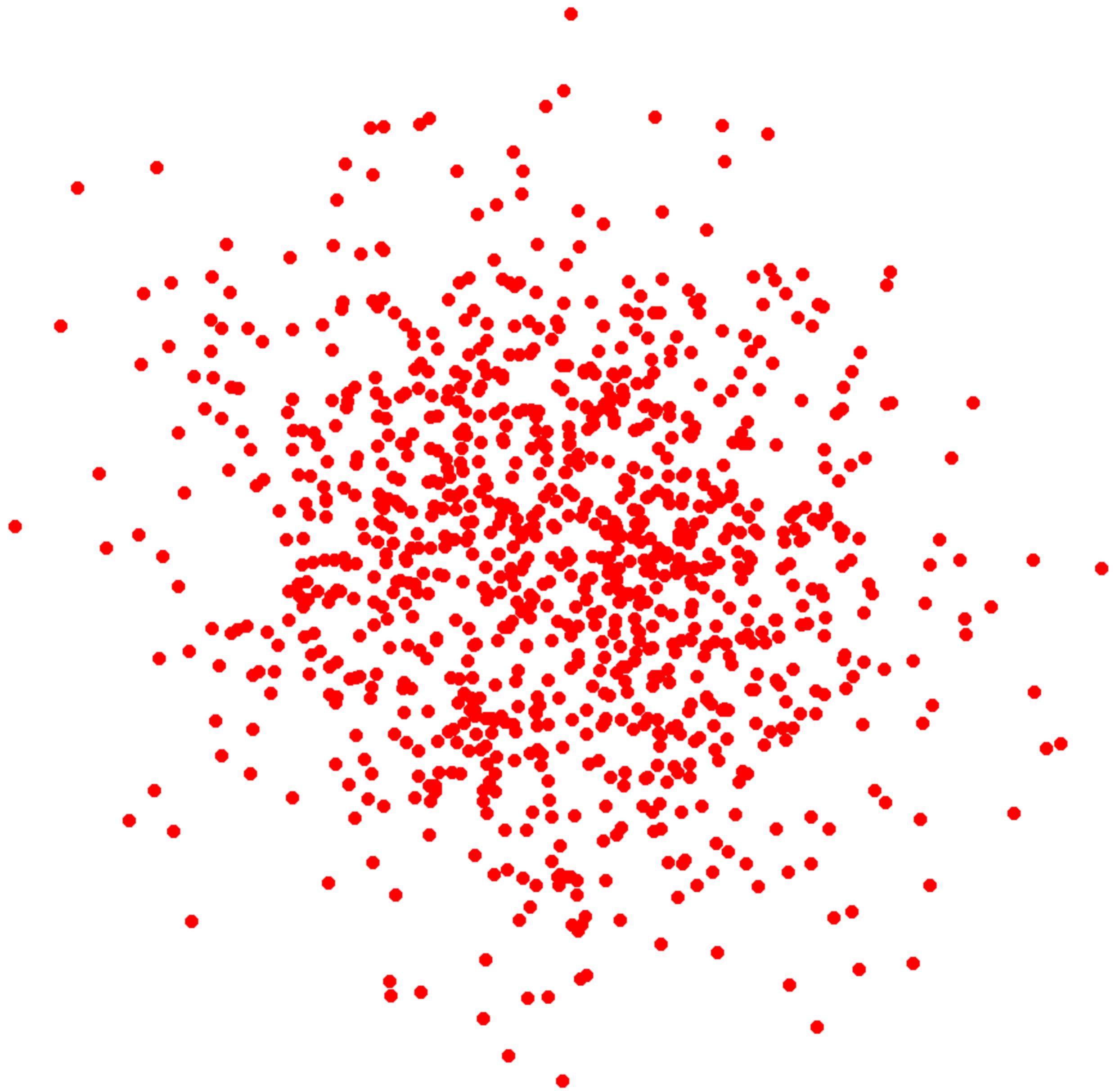
-0.25



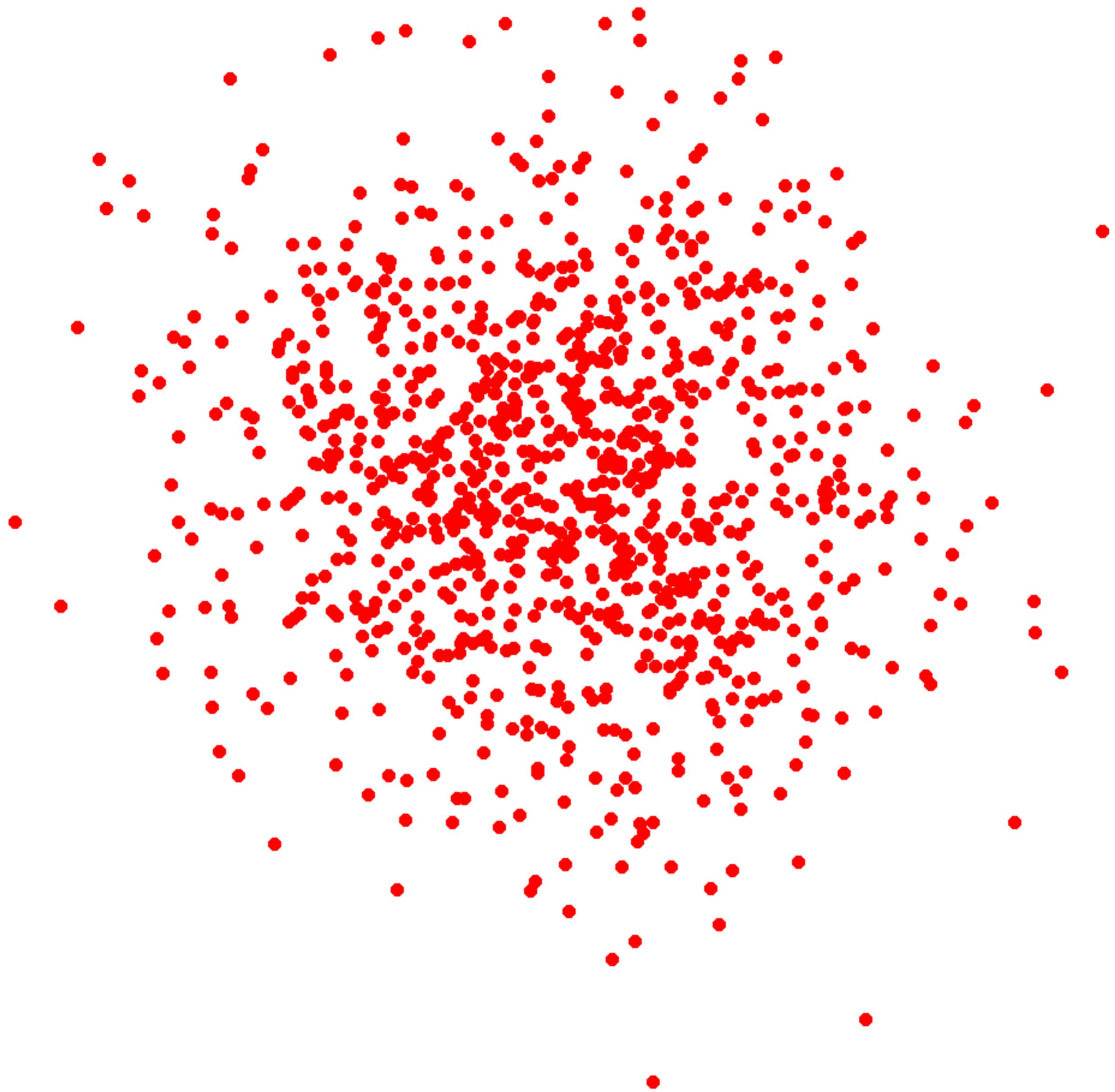
-0.14



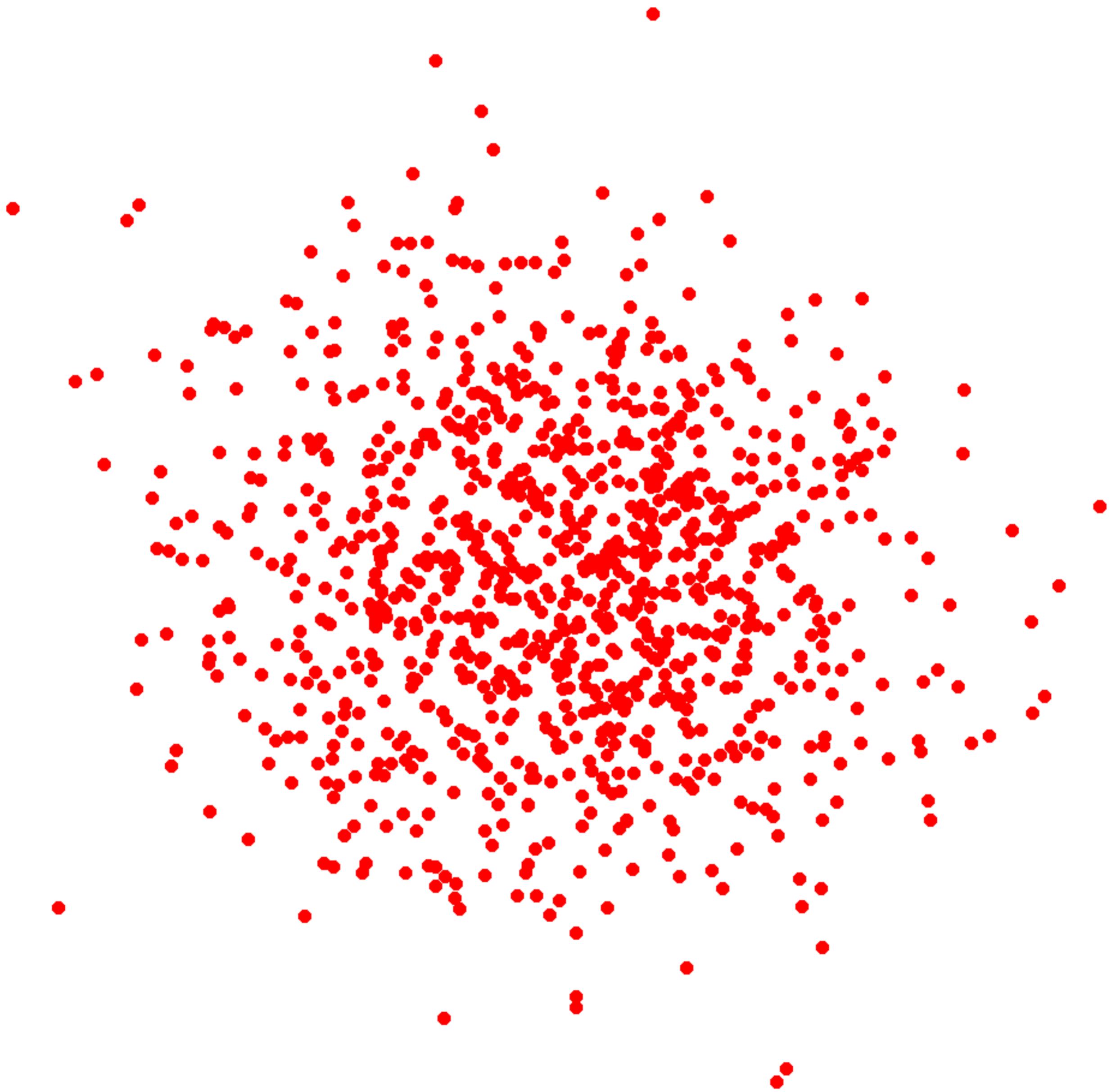
-0.12



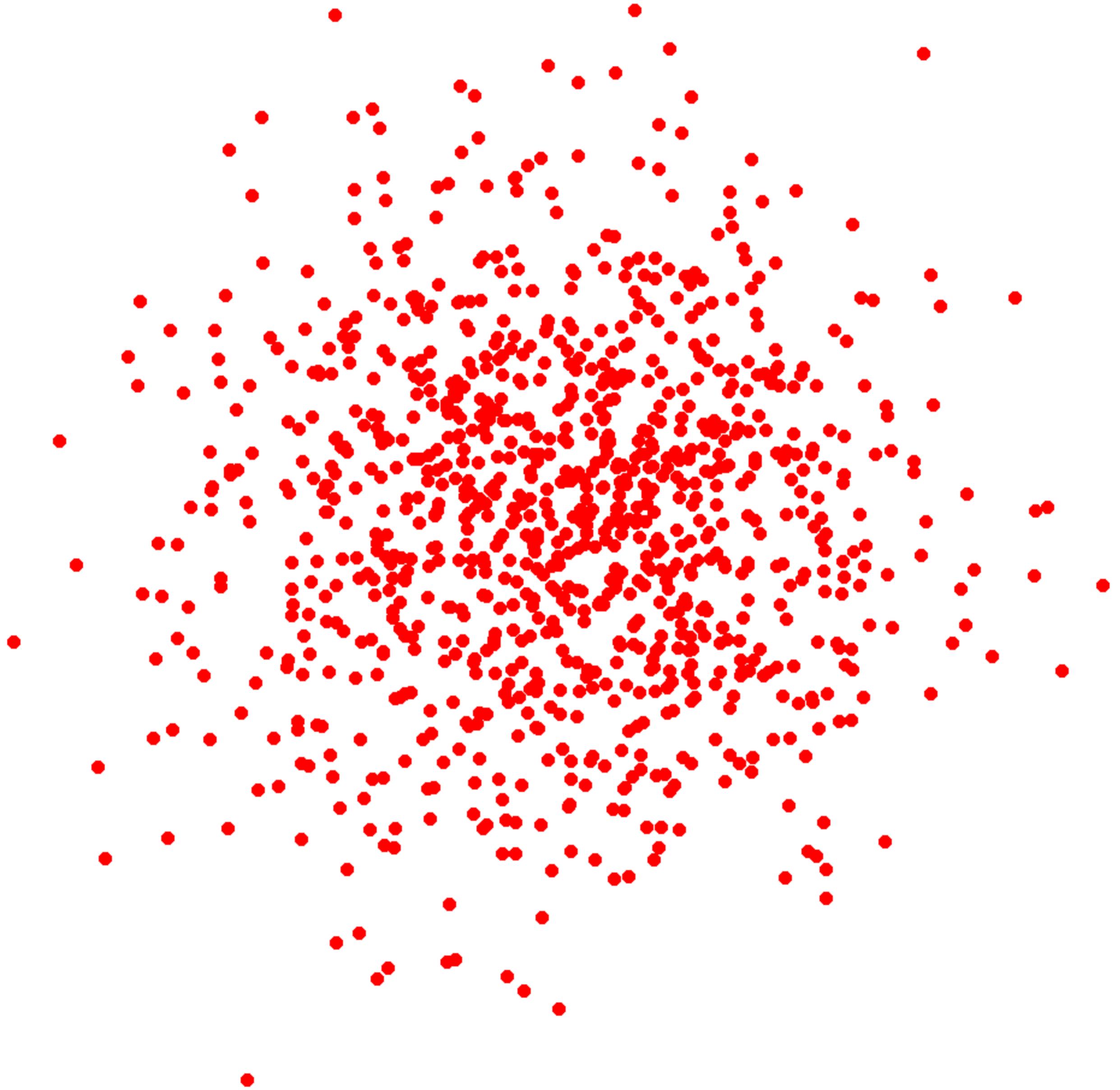
-0.09



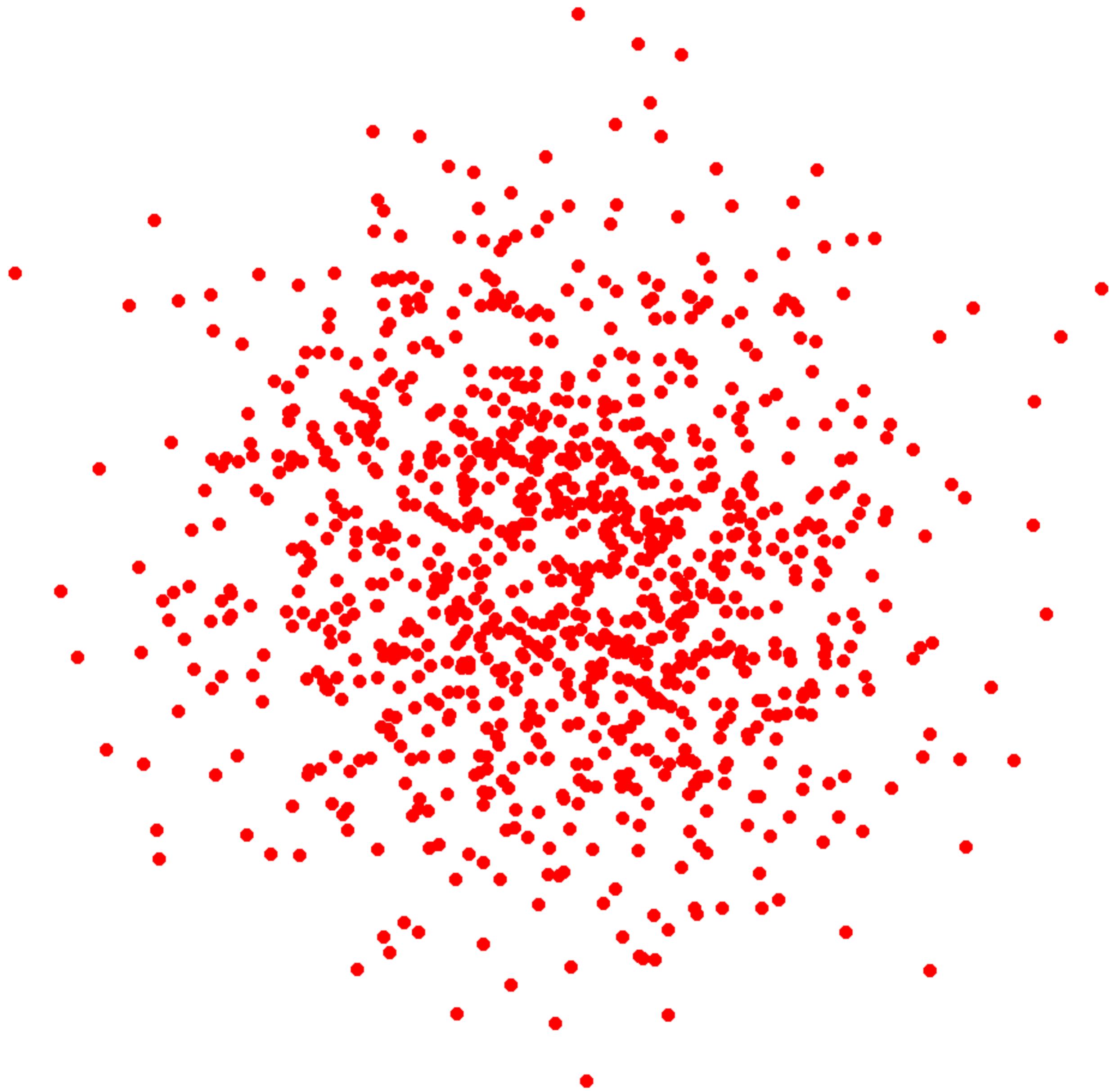
-0.07



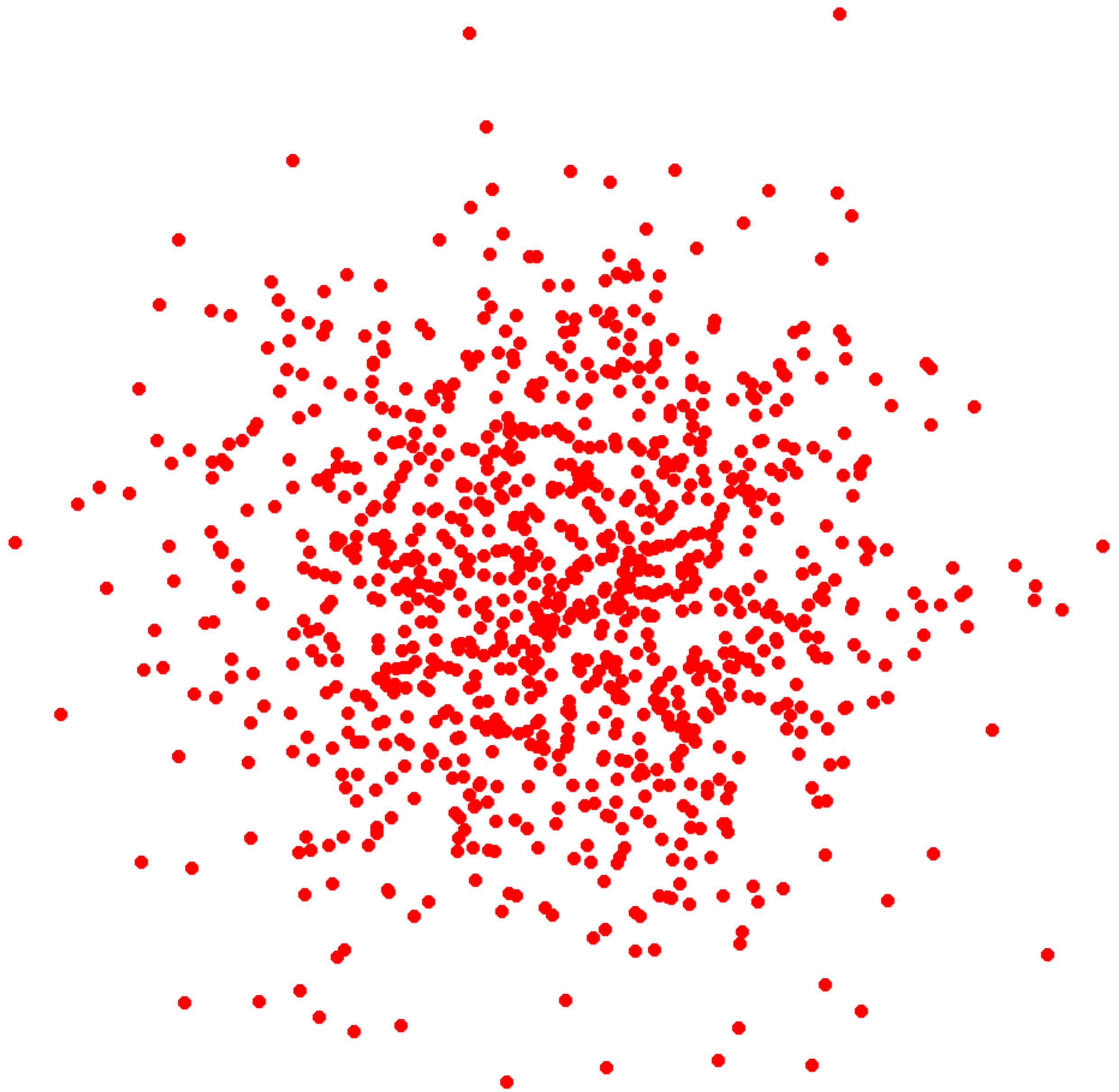
0.01



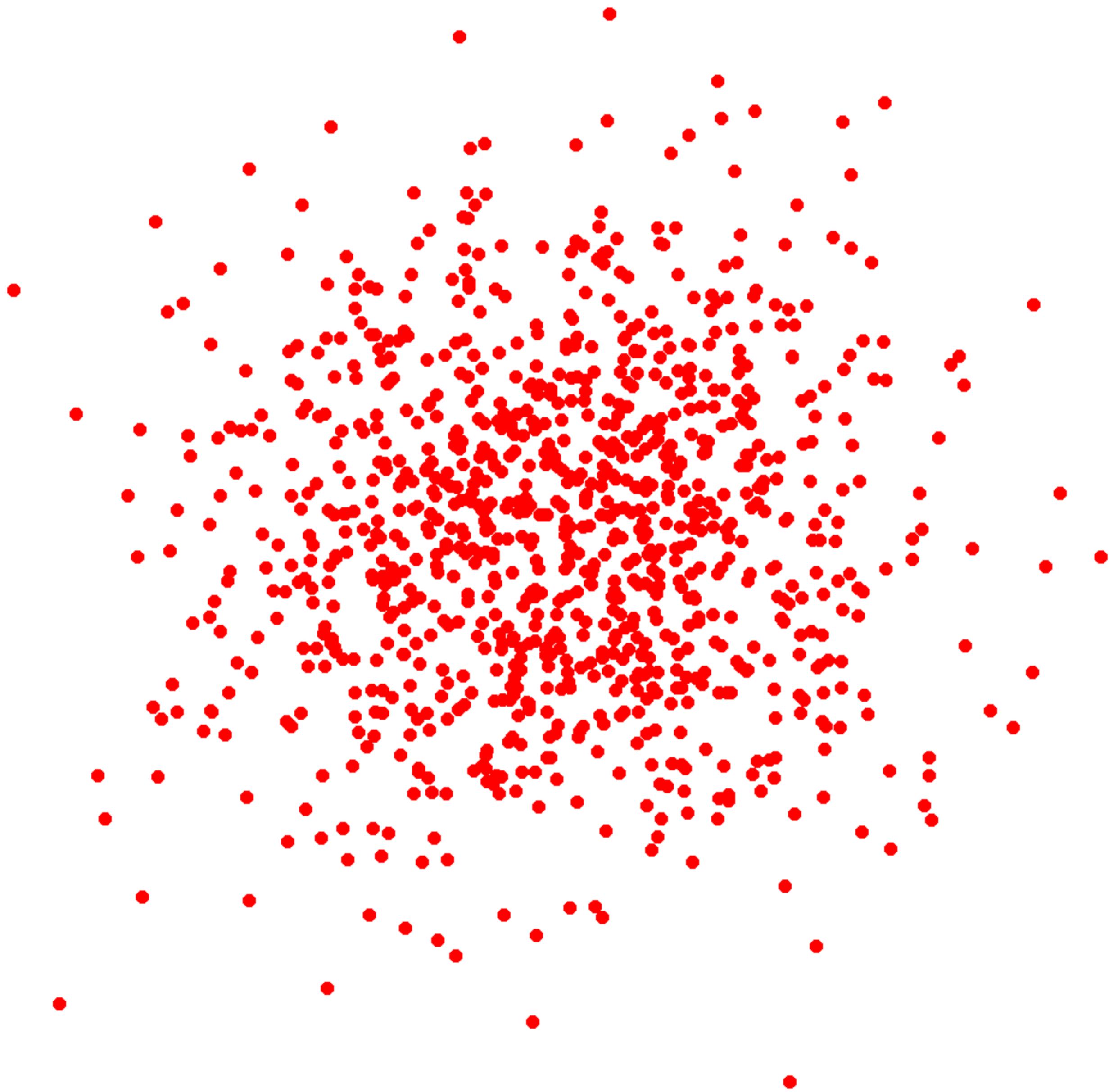
-0.05



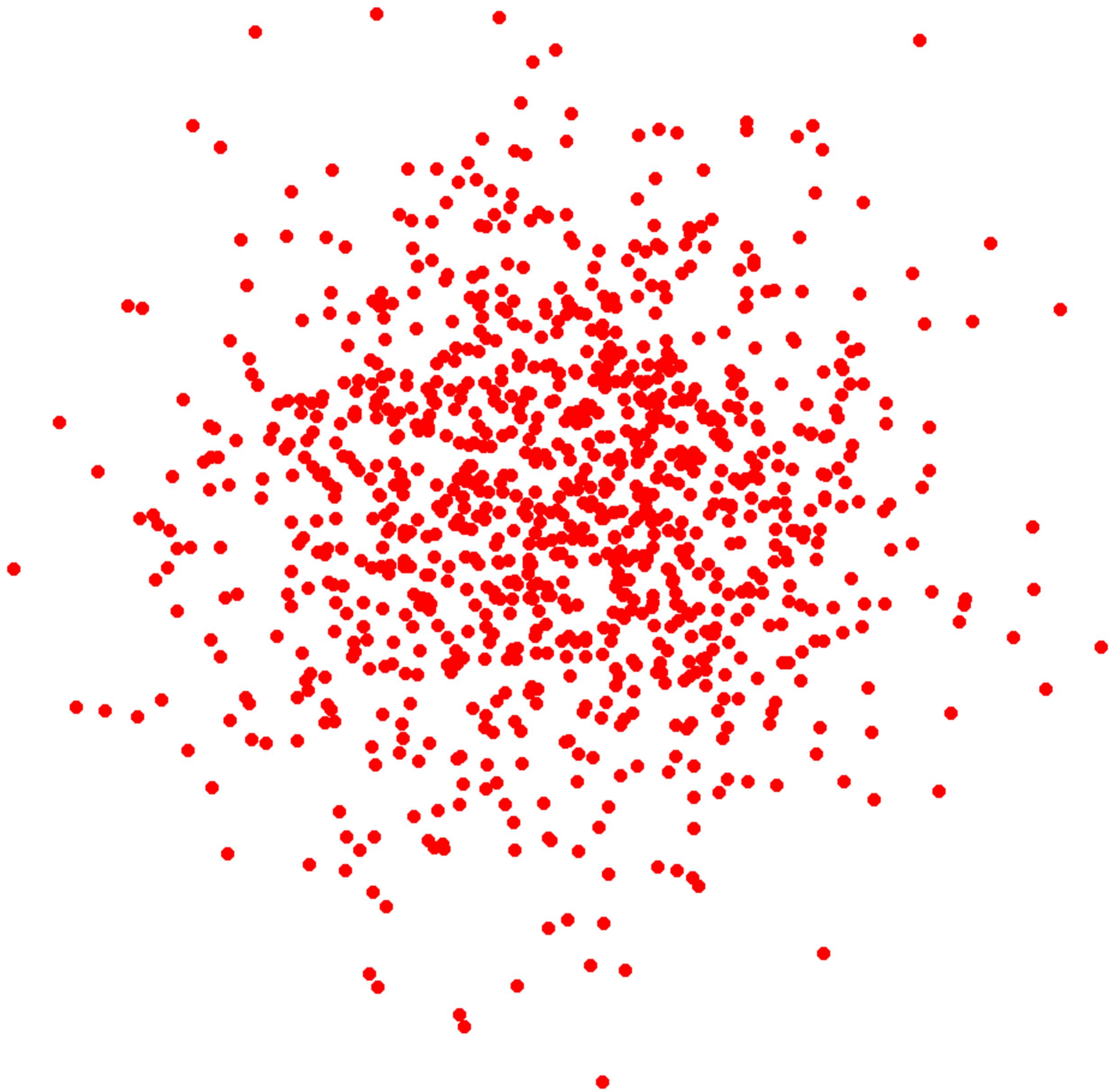
-0.01



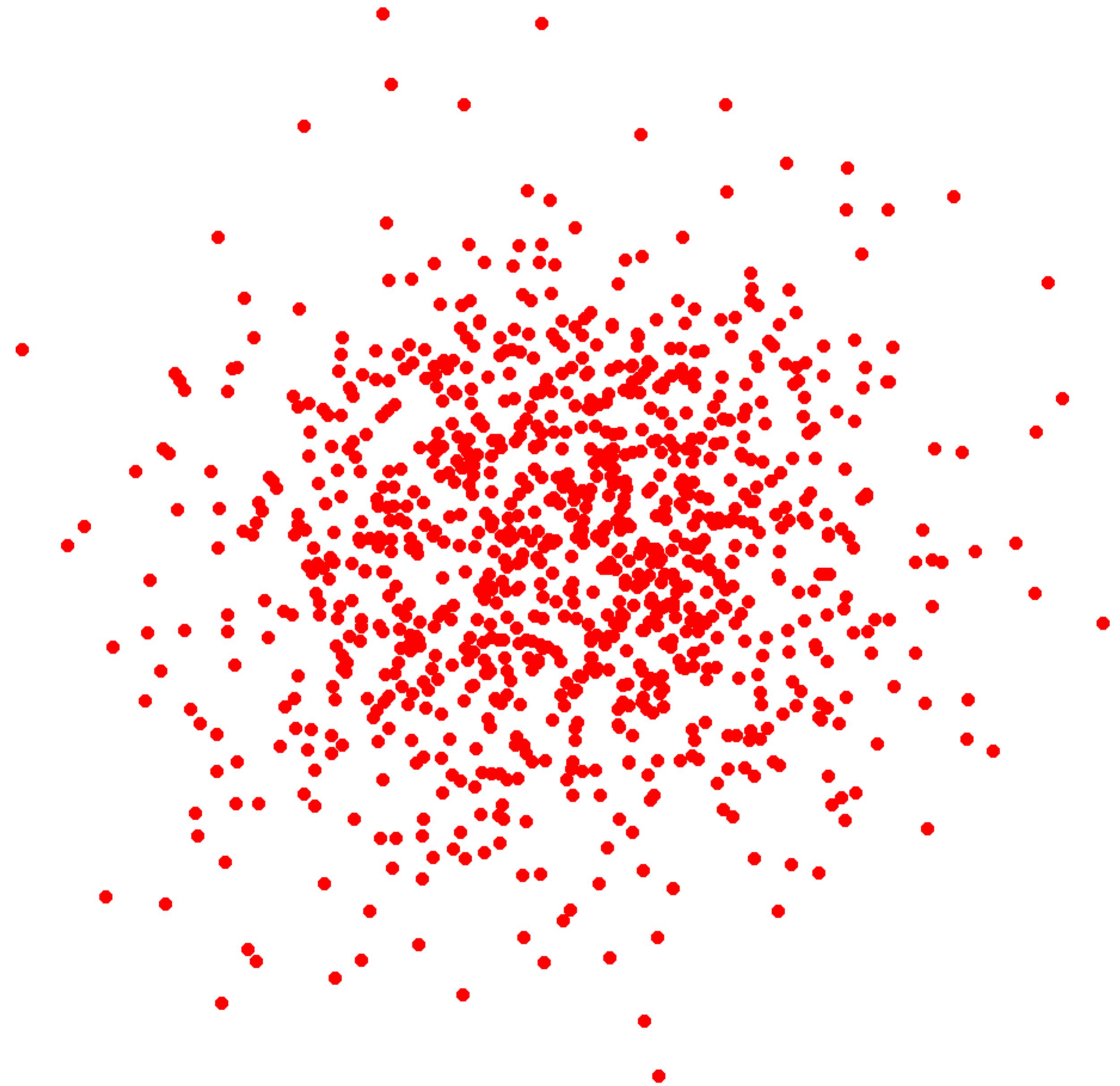
0.04



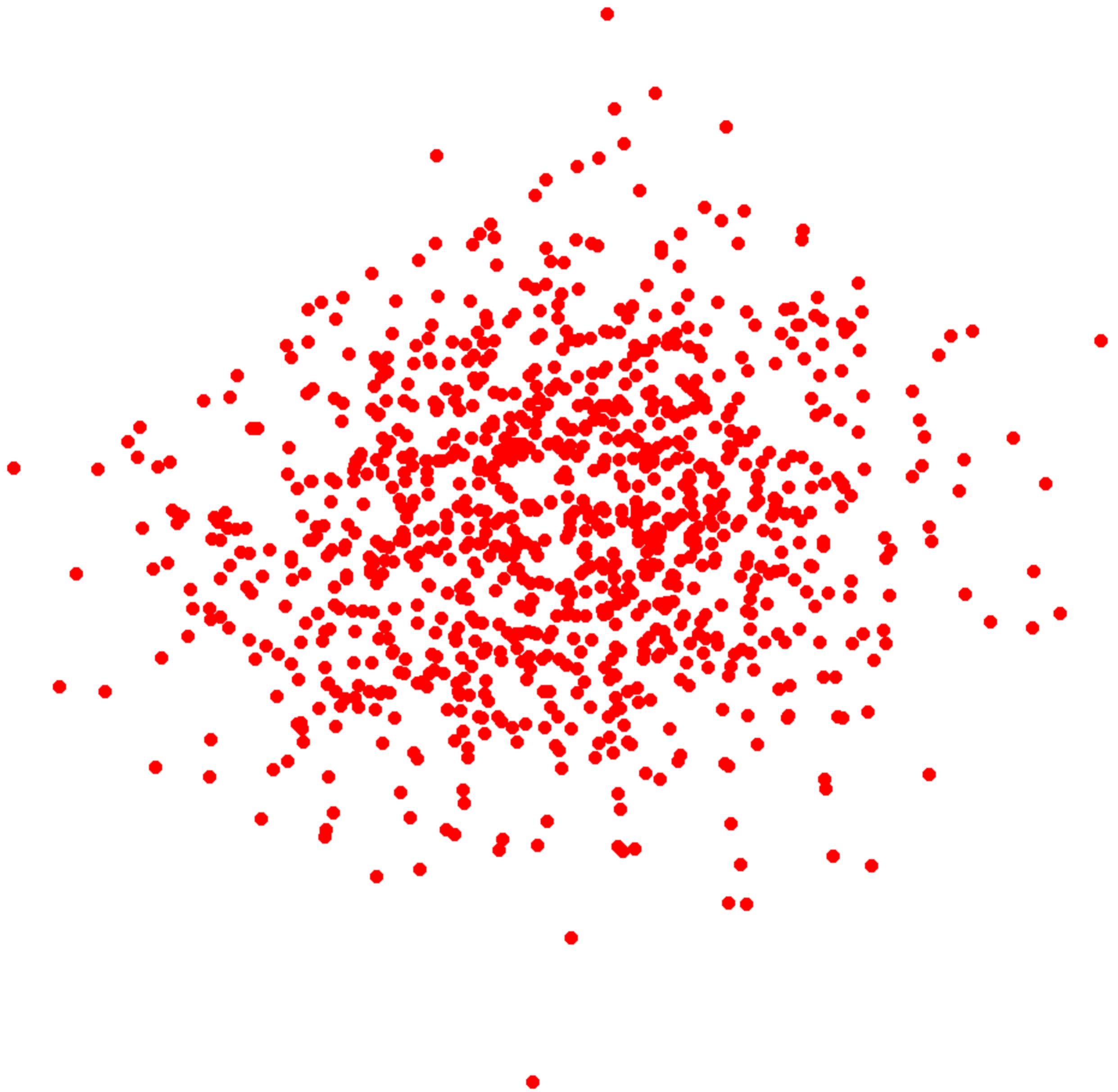
0.03



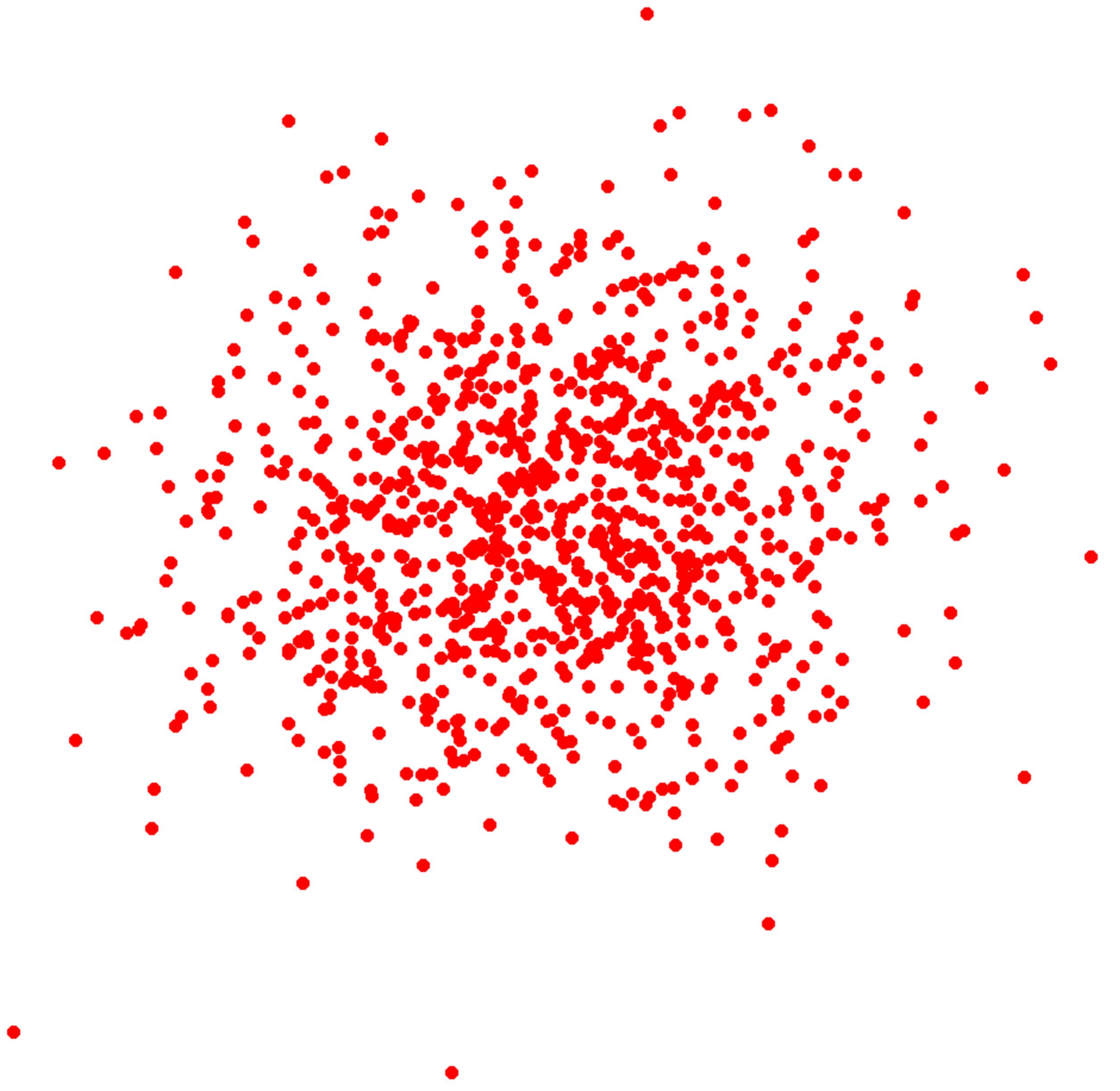
0.07



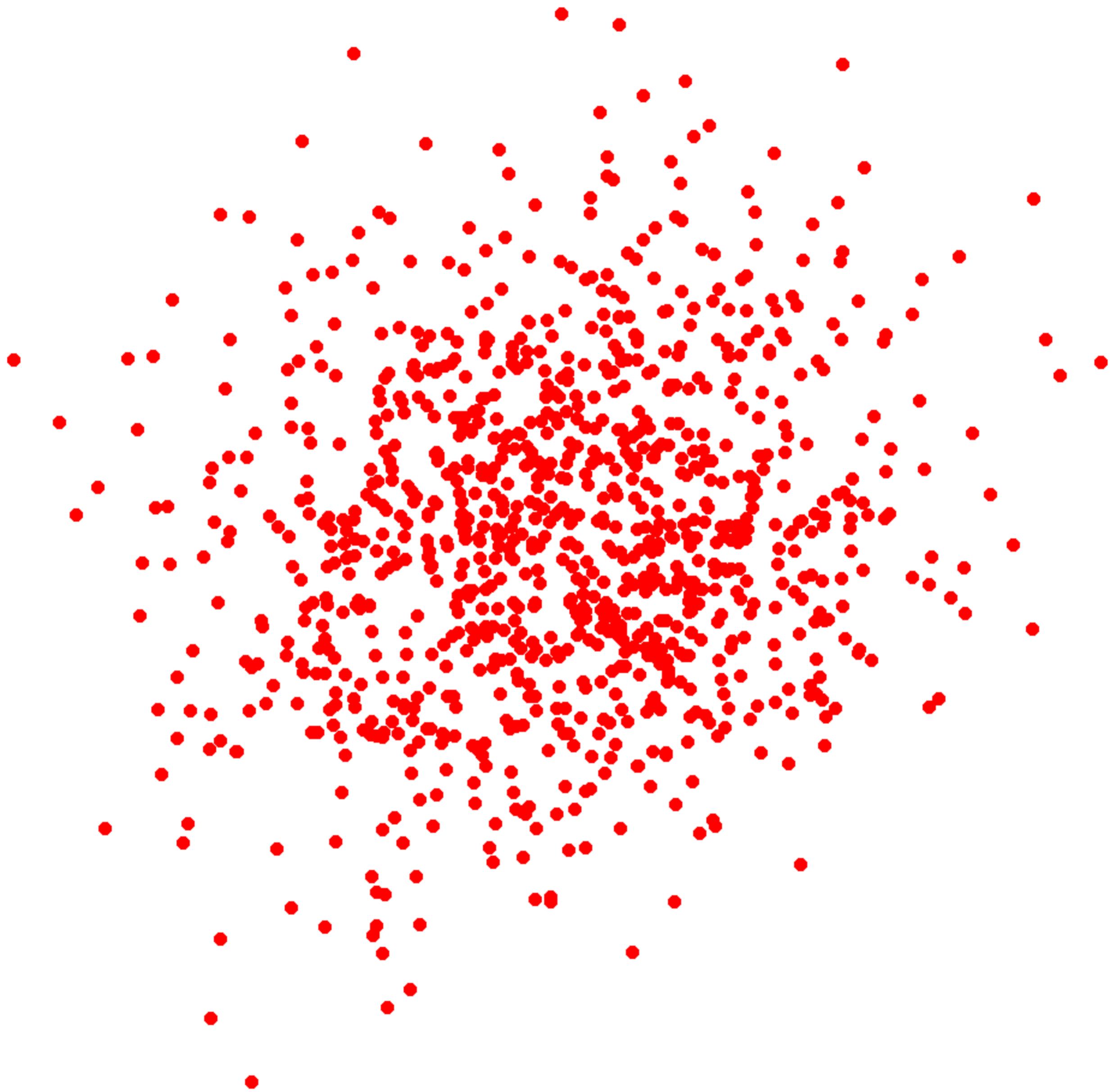
0.11



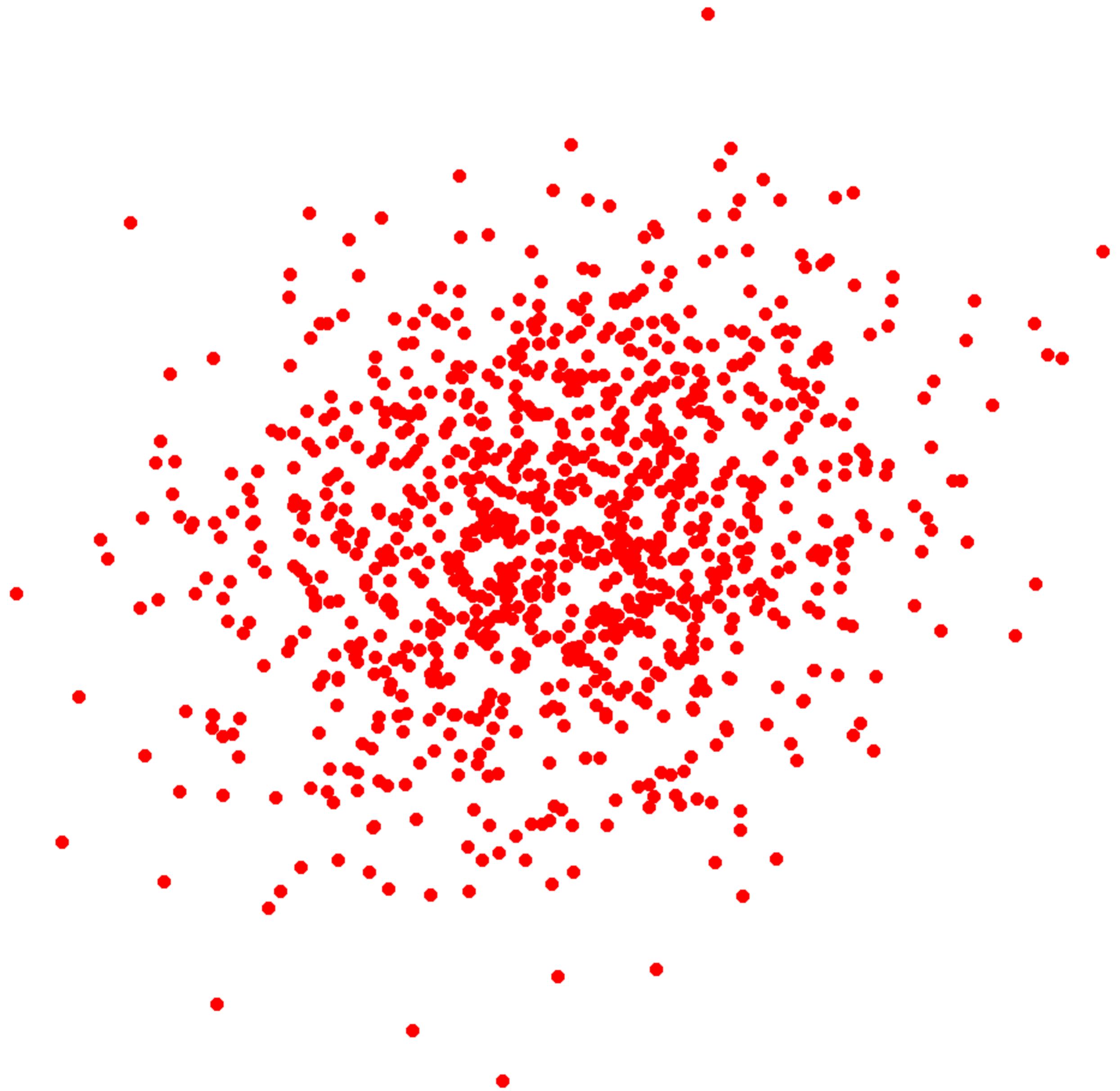
0.1



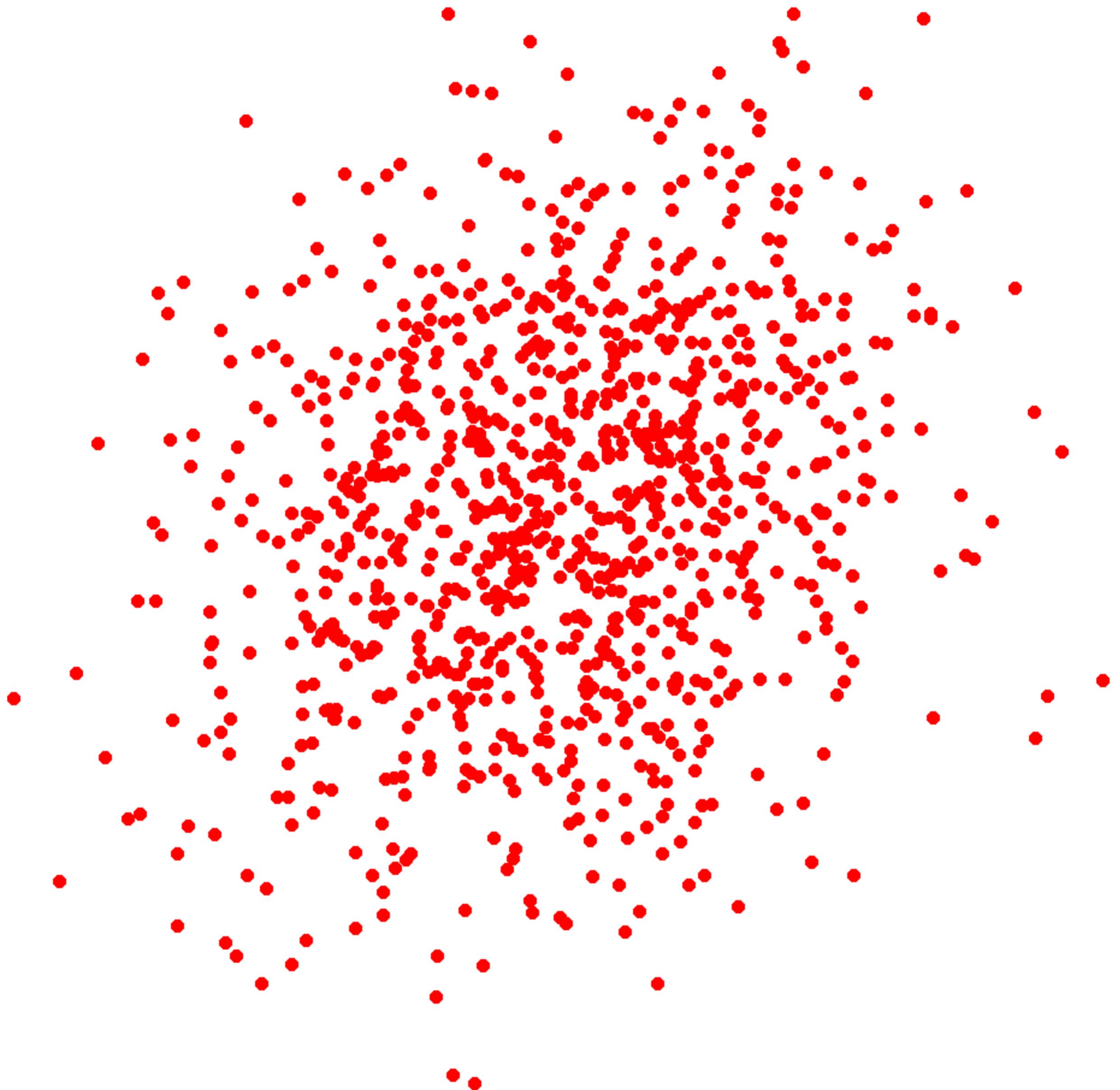
0.15



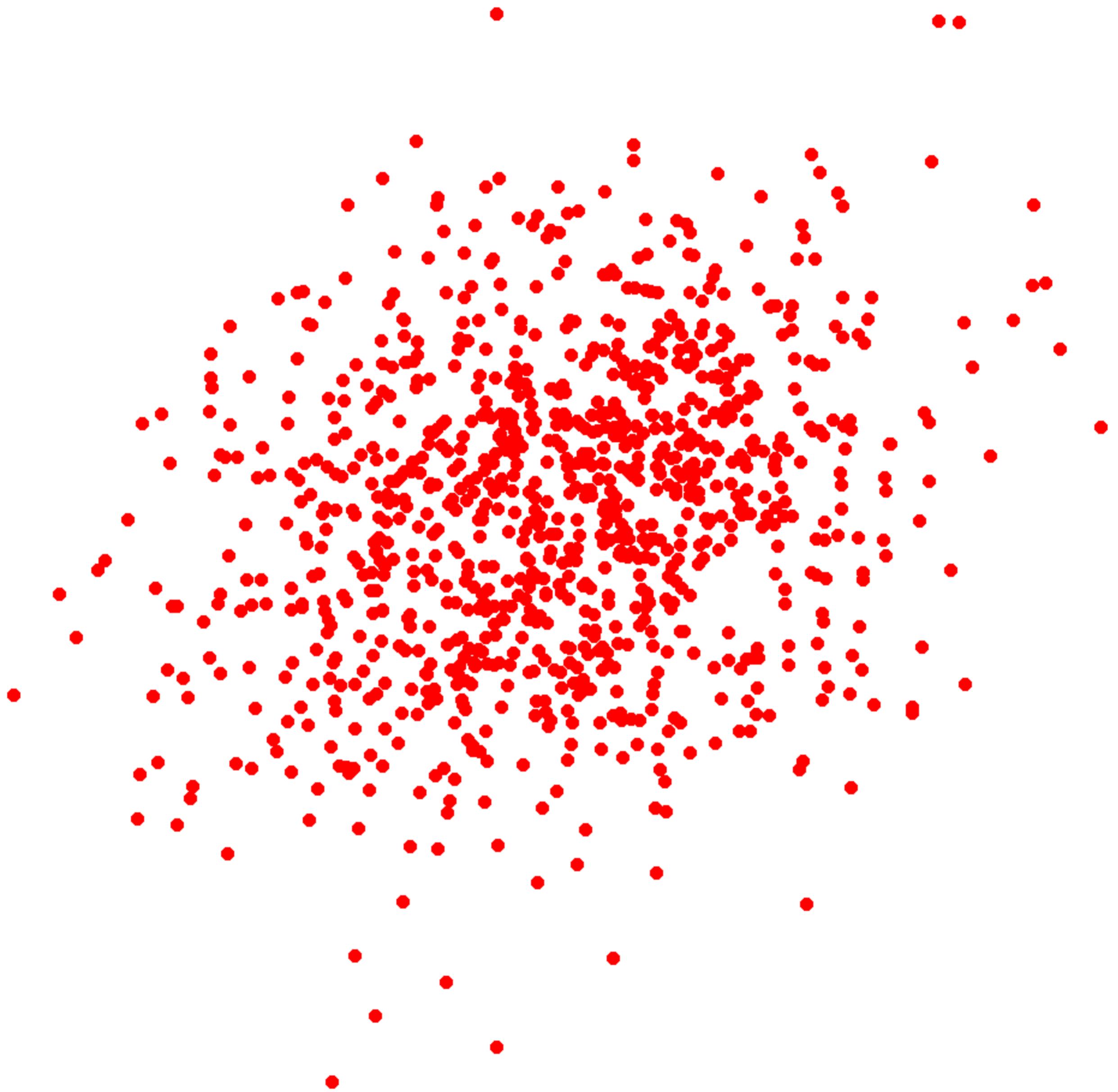
0.2



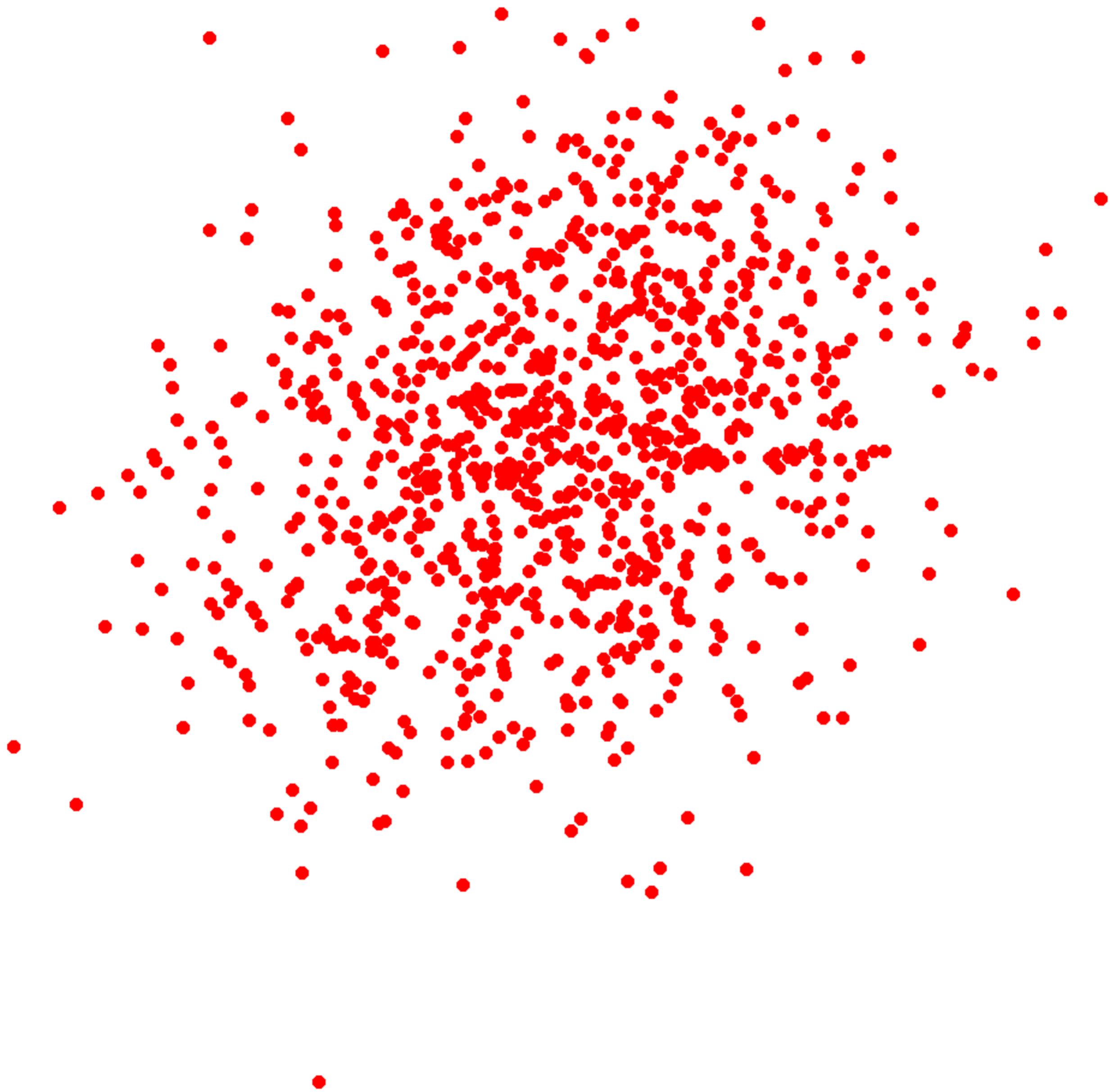
0.23



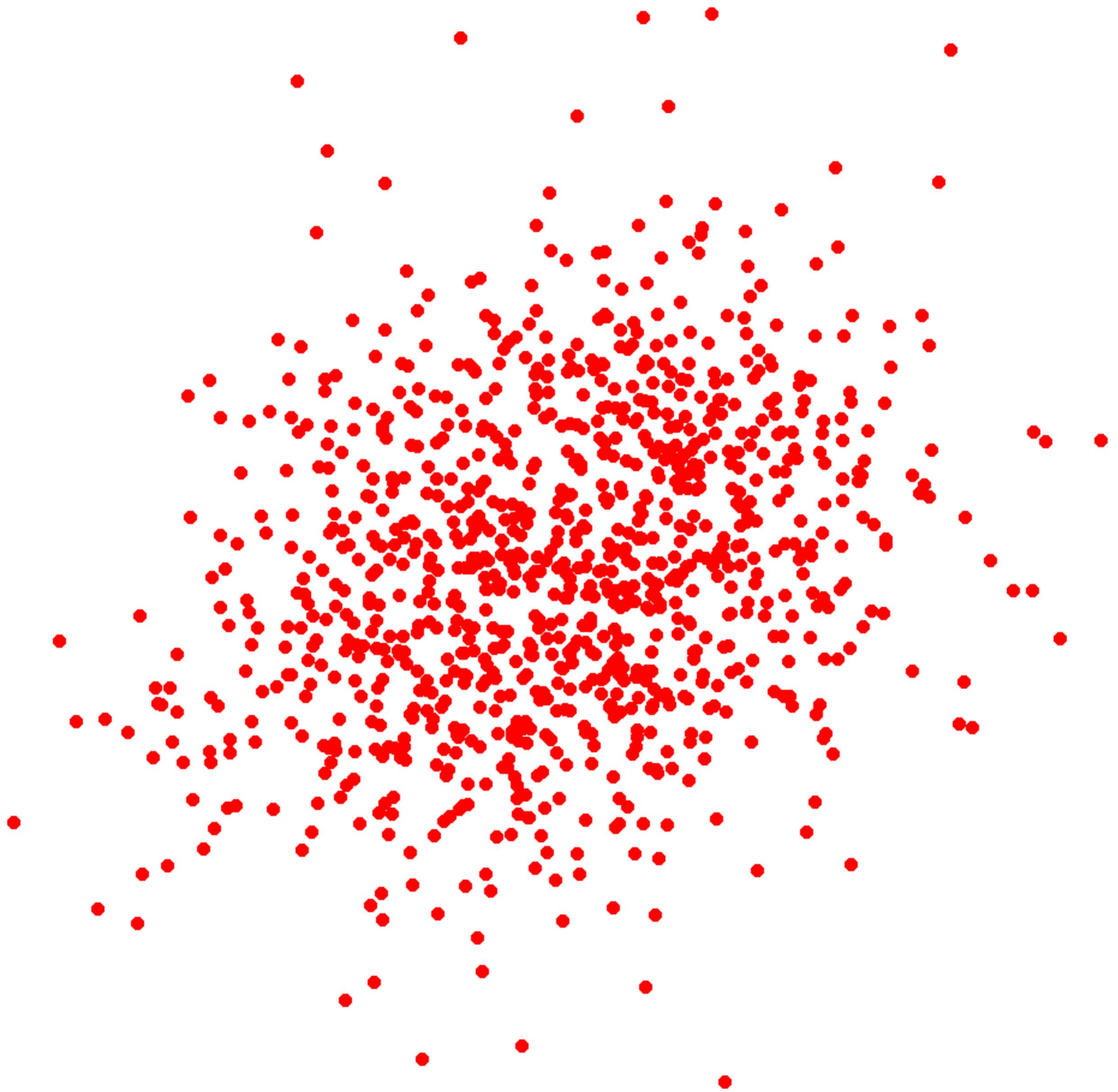
0.24



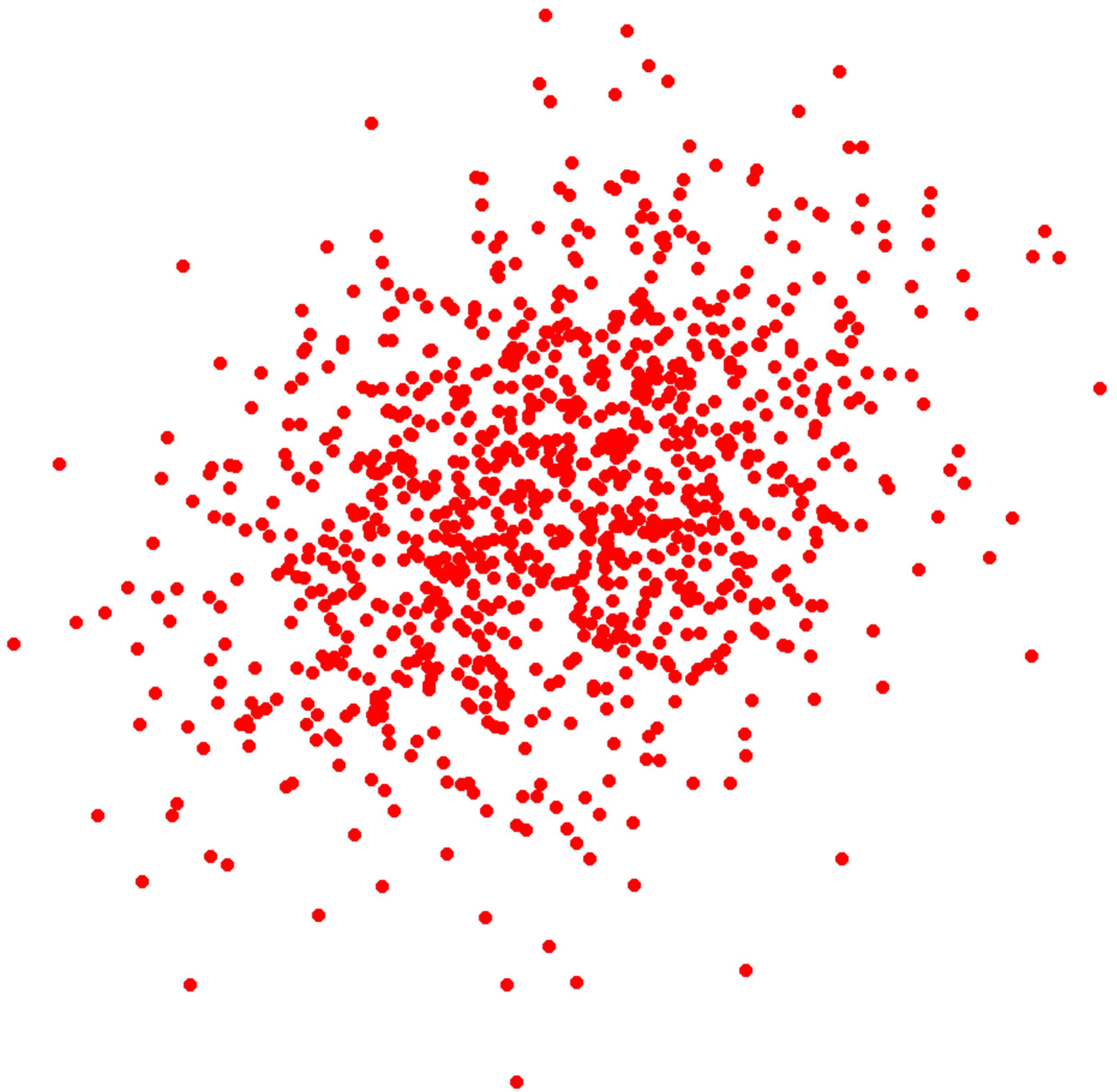
0.25



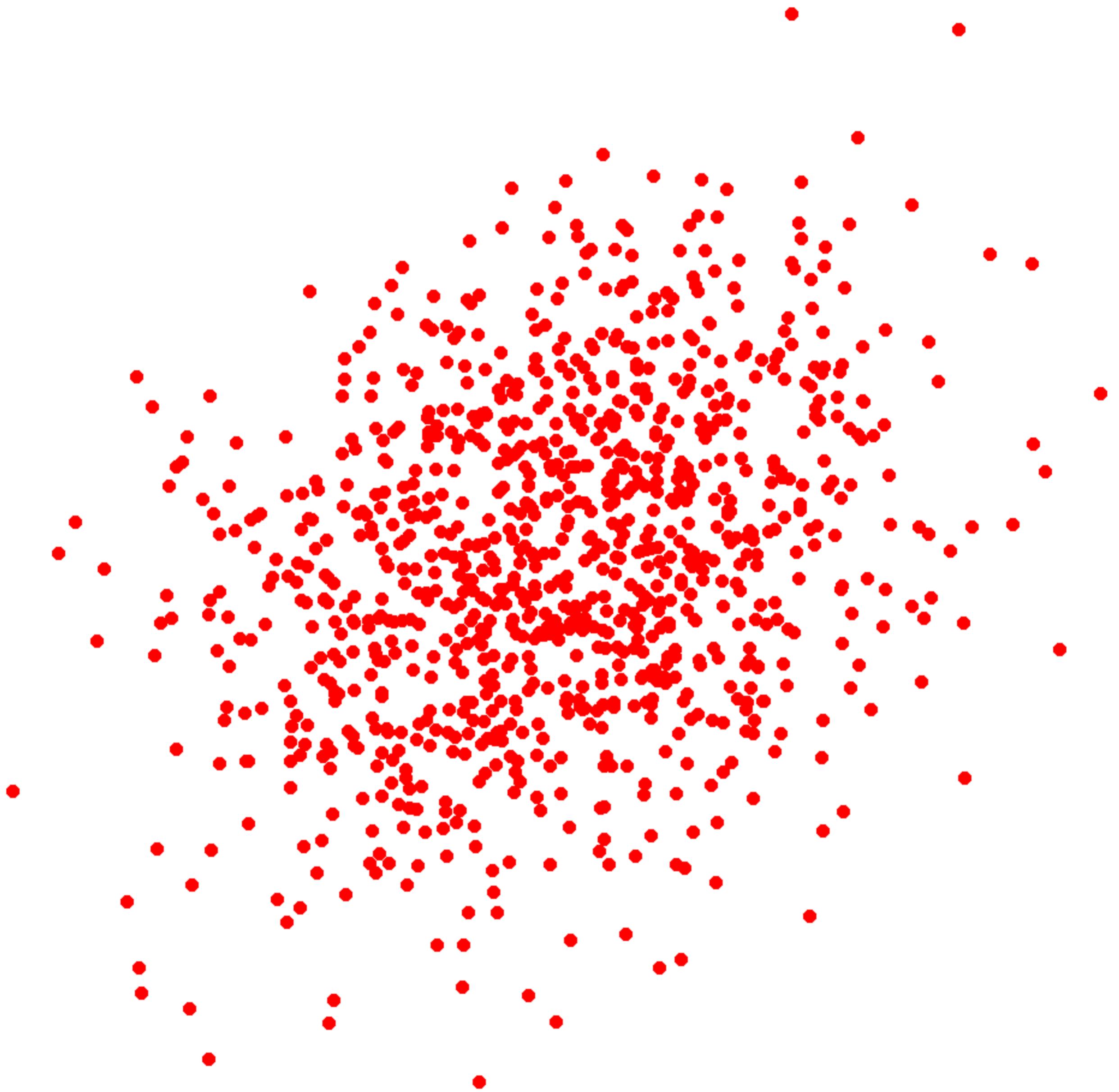
0.25



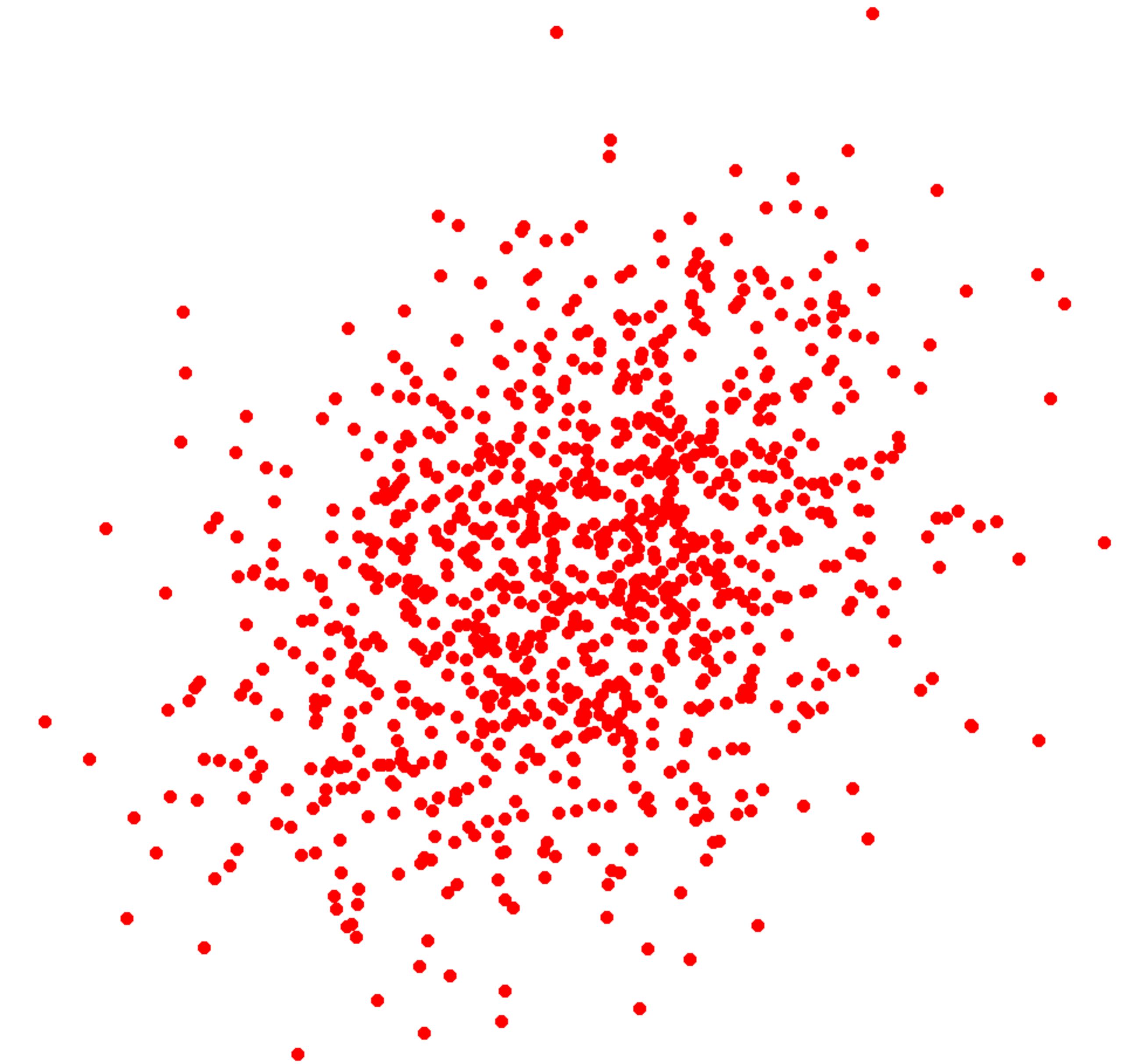
0.31



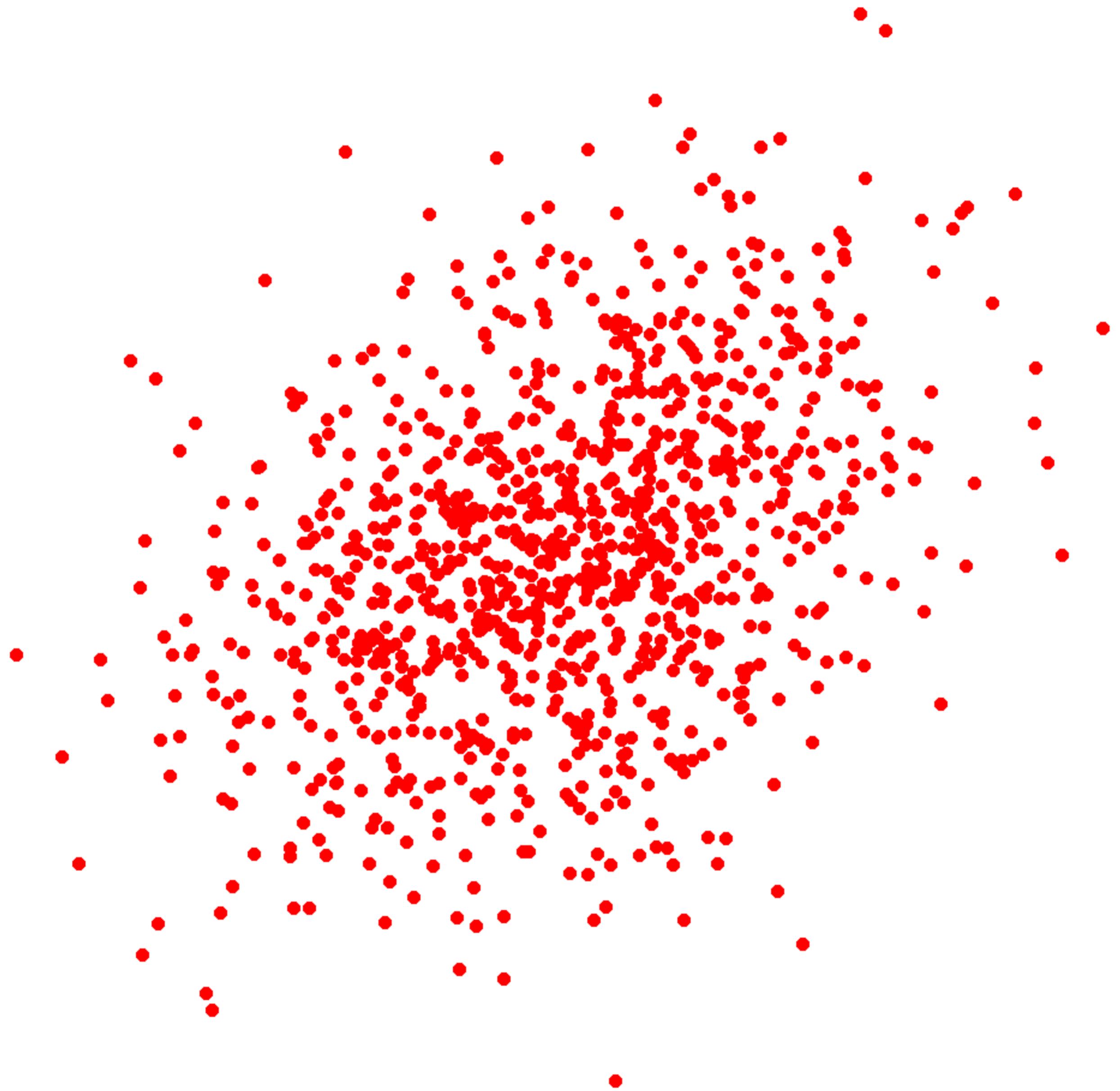
0.29



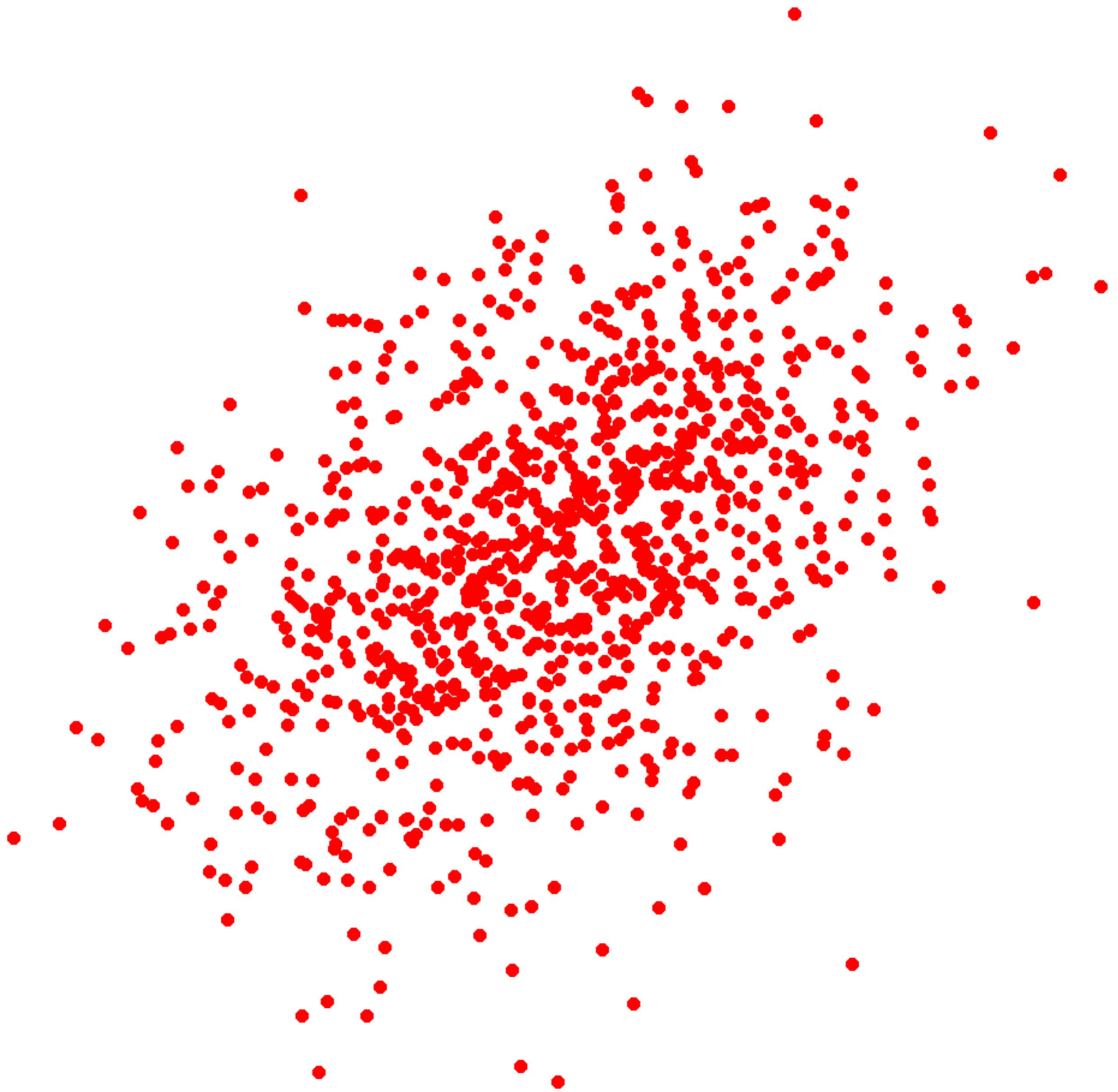
0.33



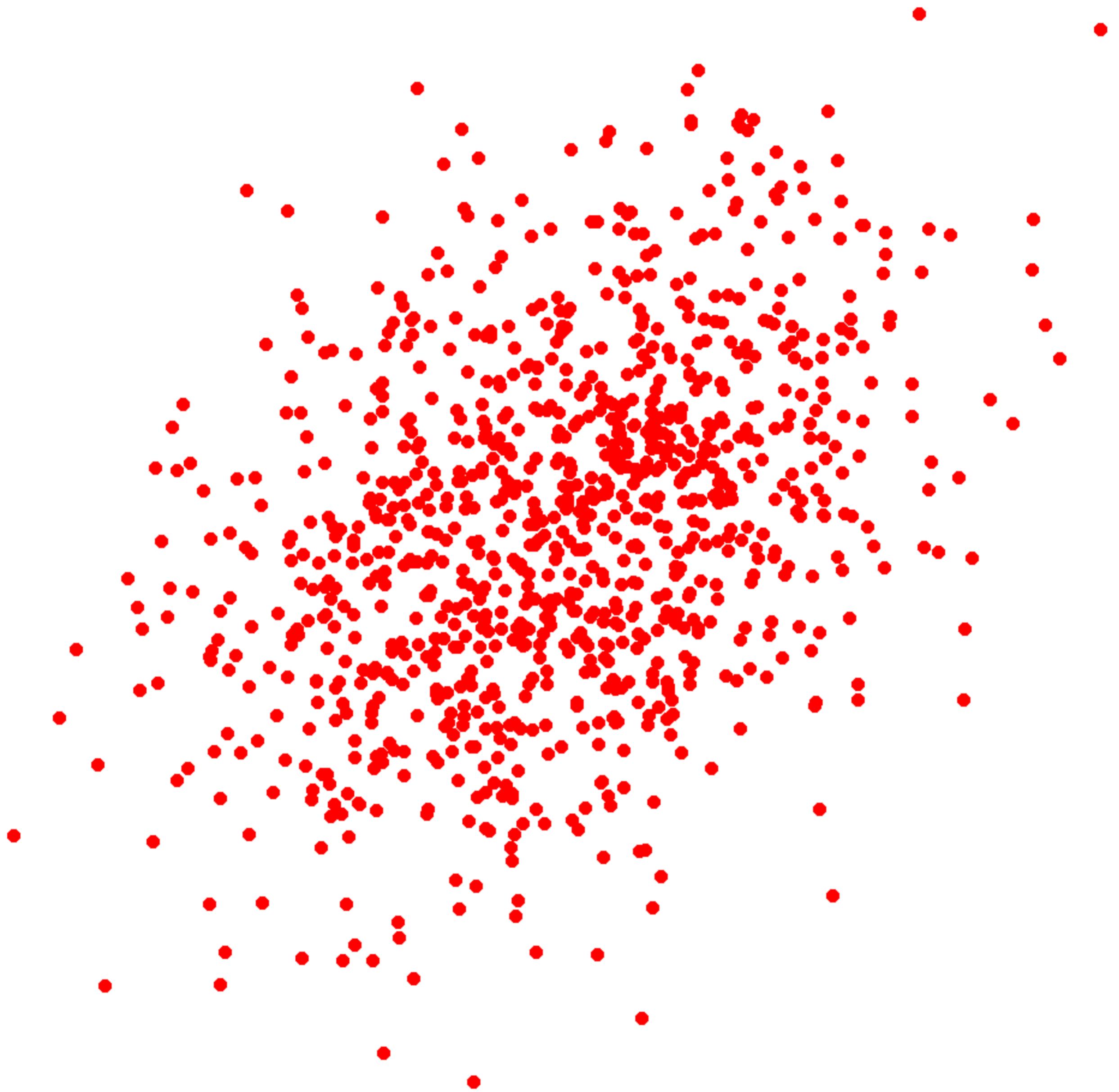
0.37



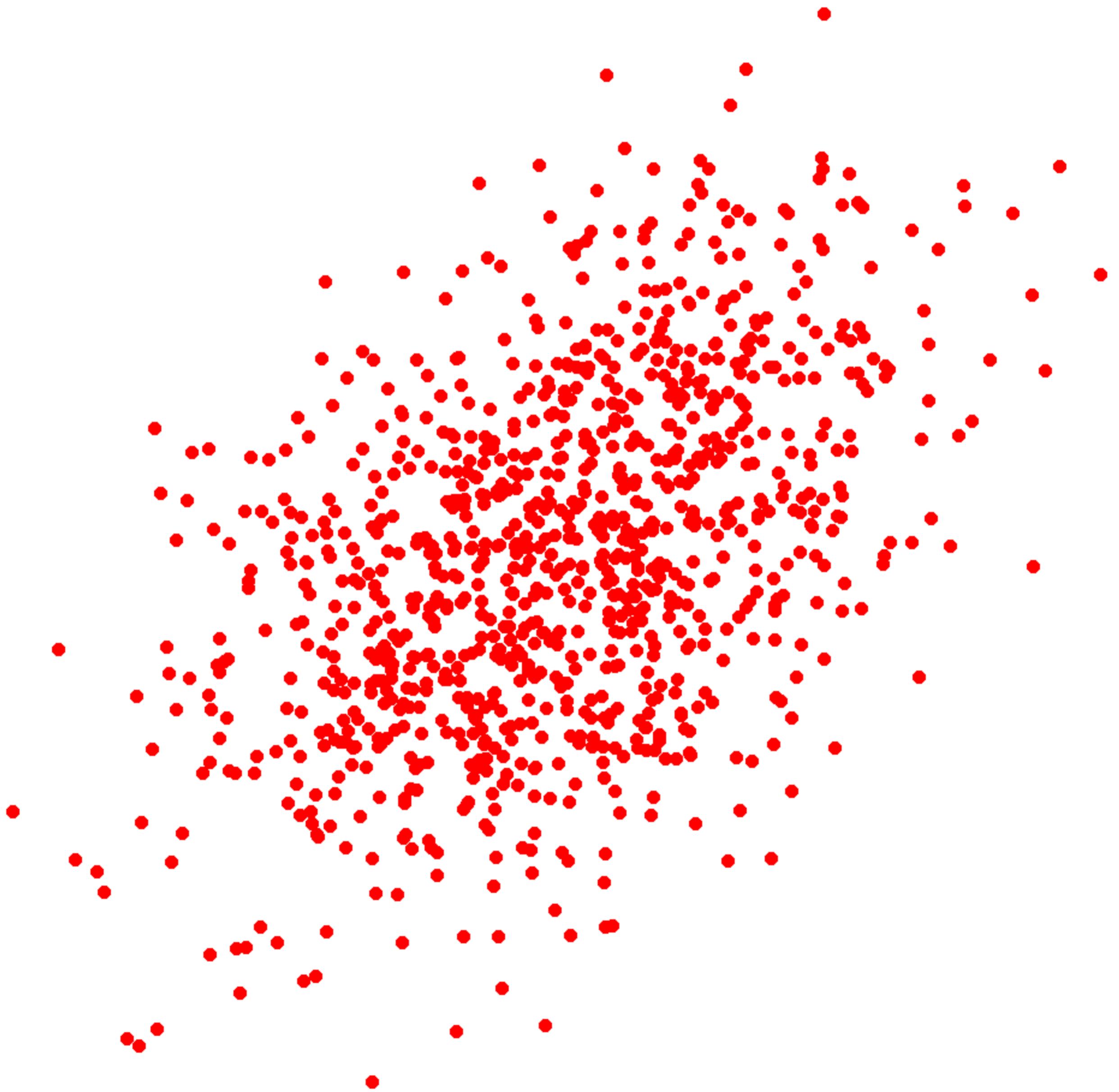
0.43



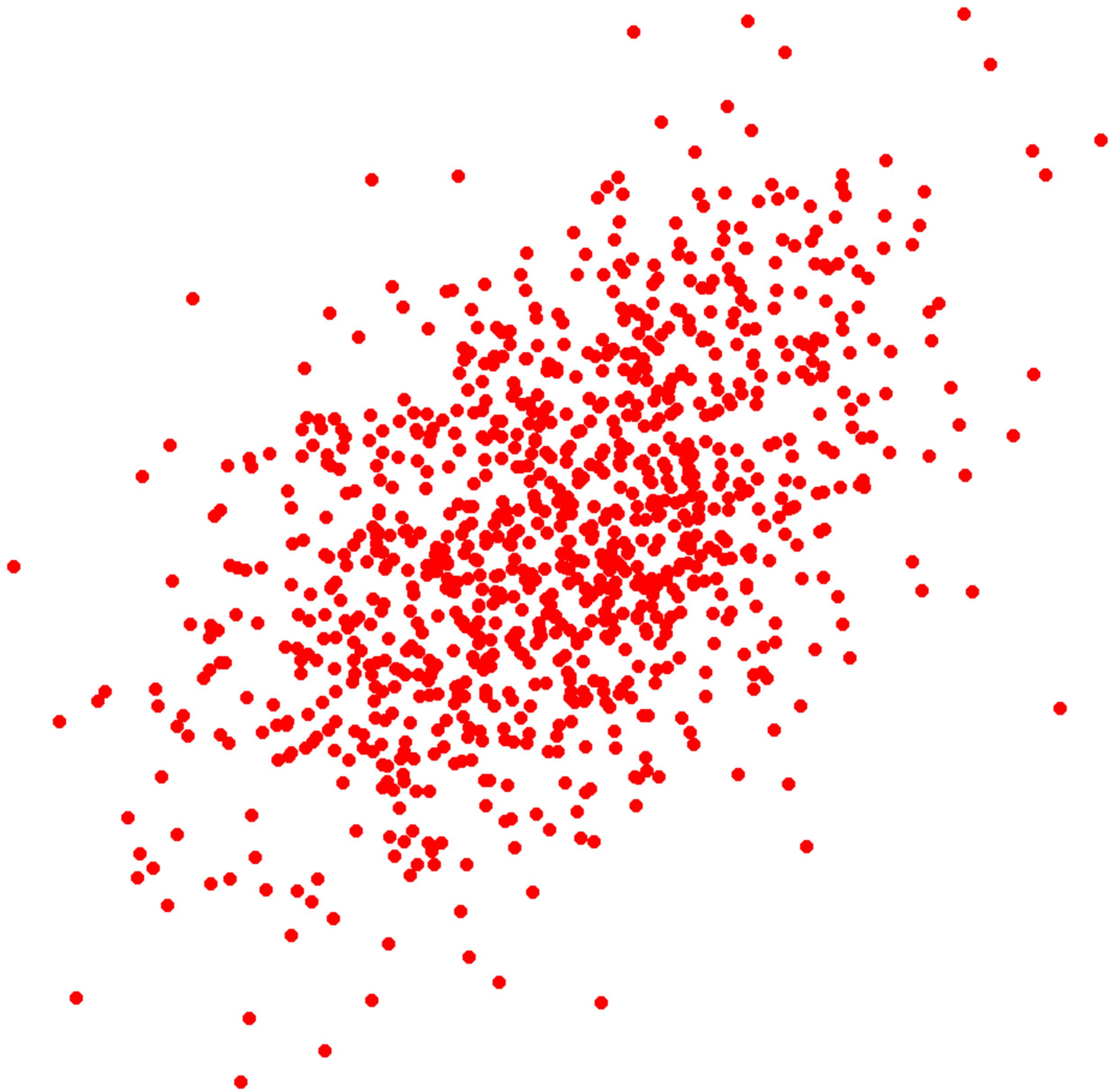
0.38



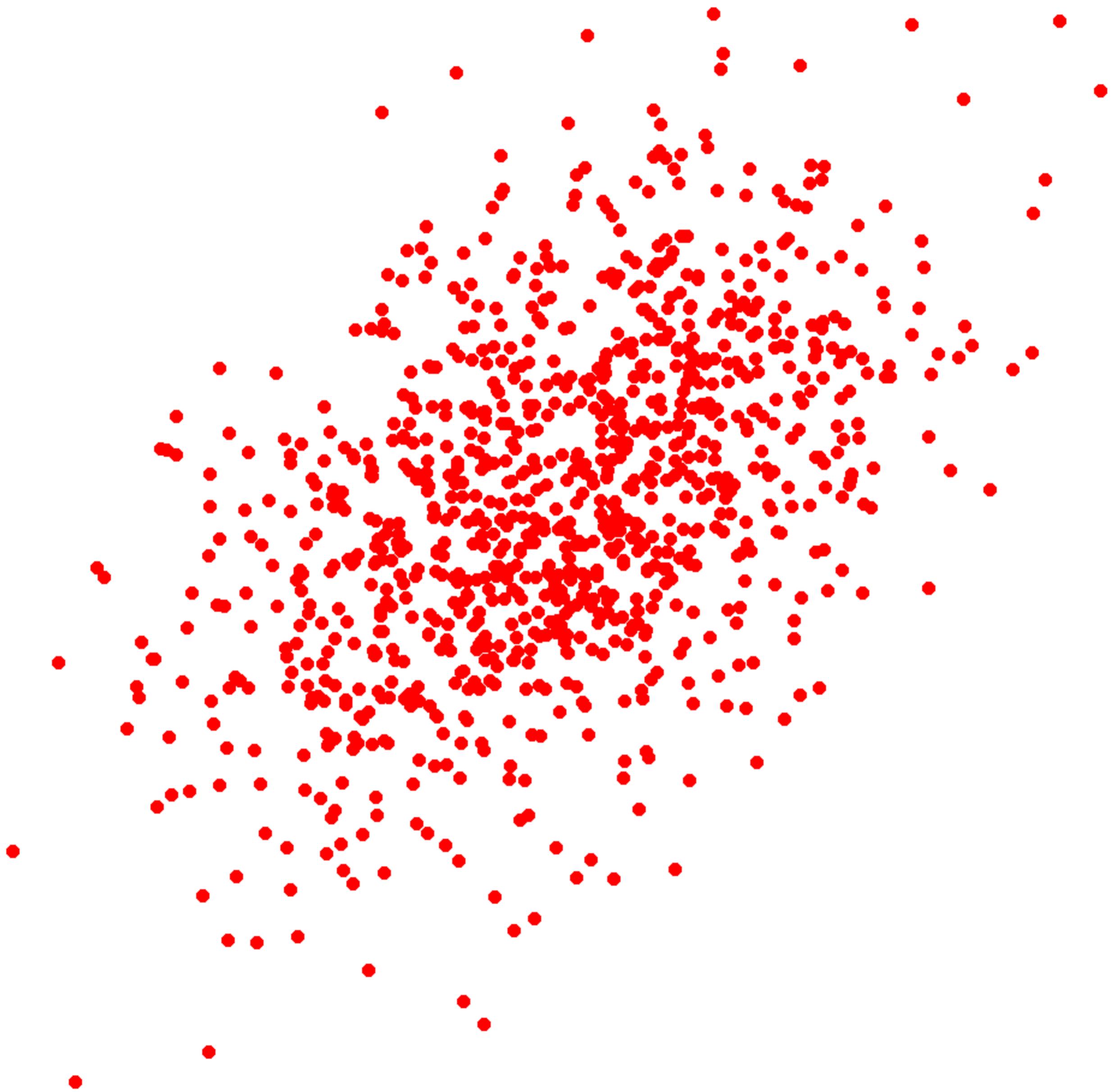
0.47



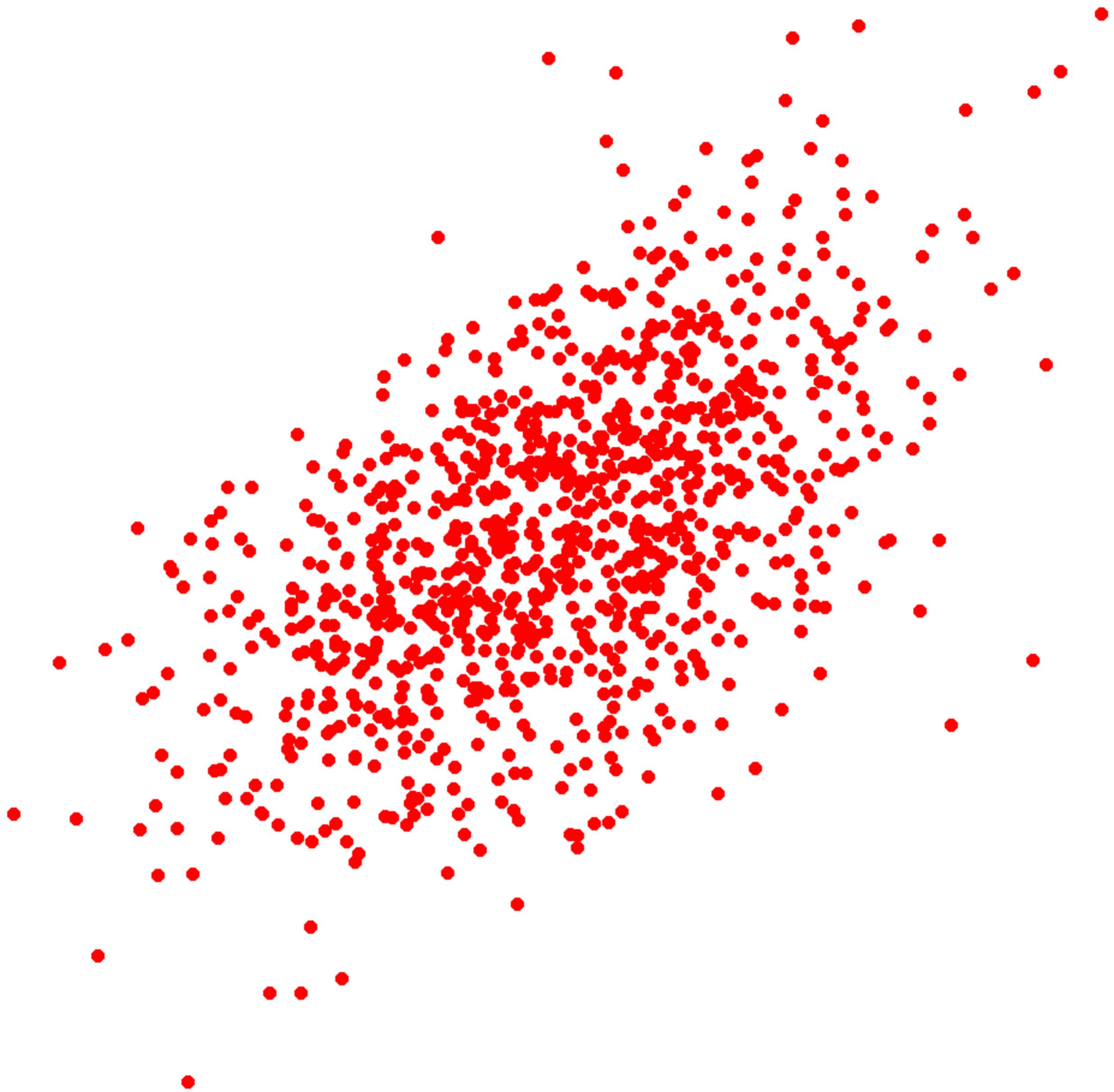
0.49



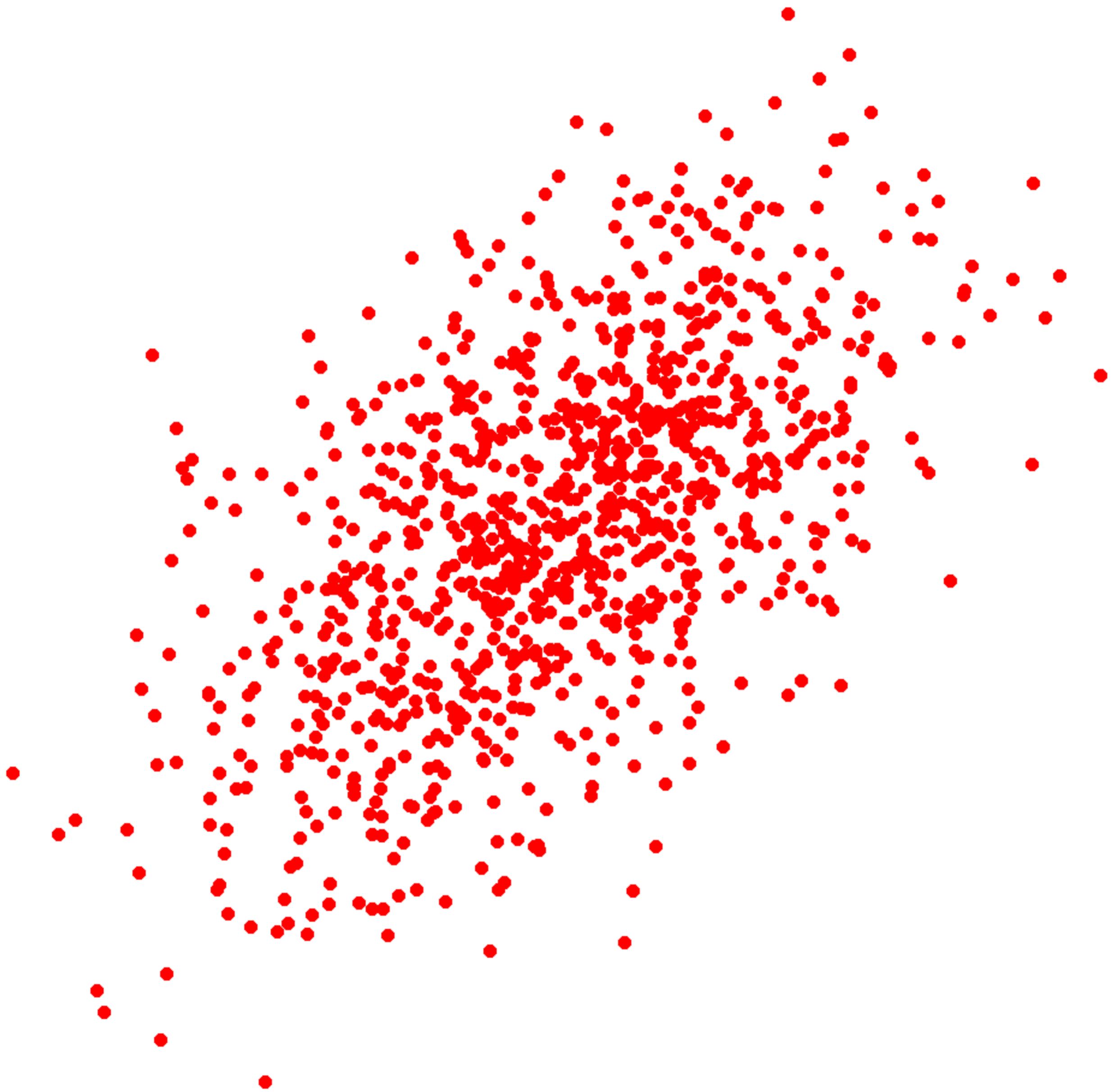
0.47



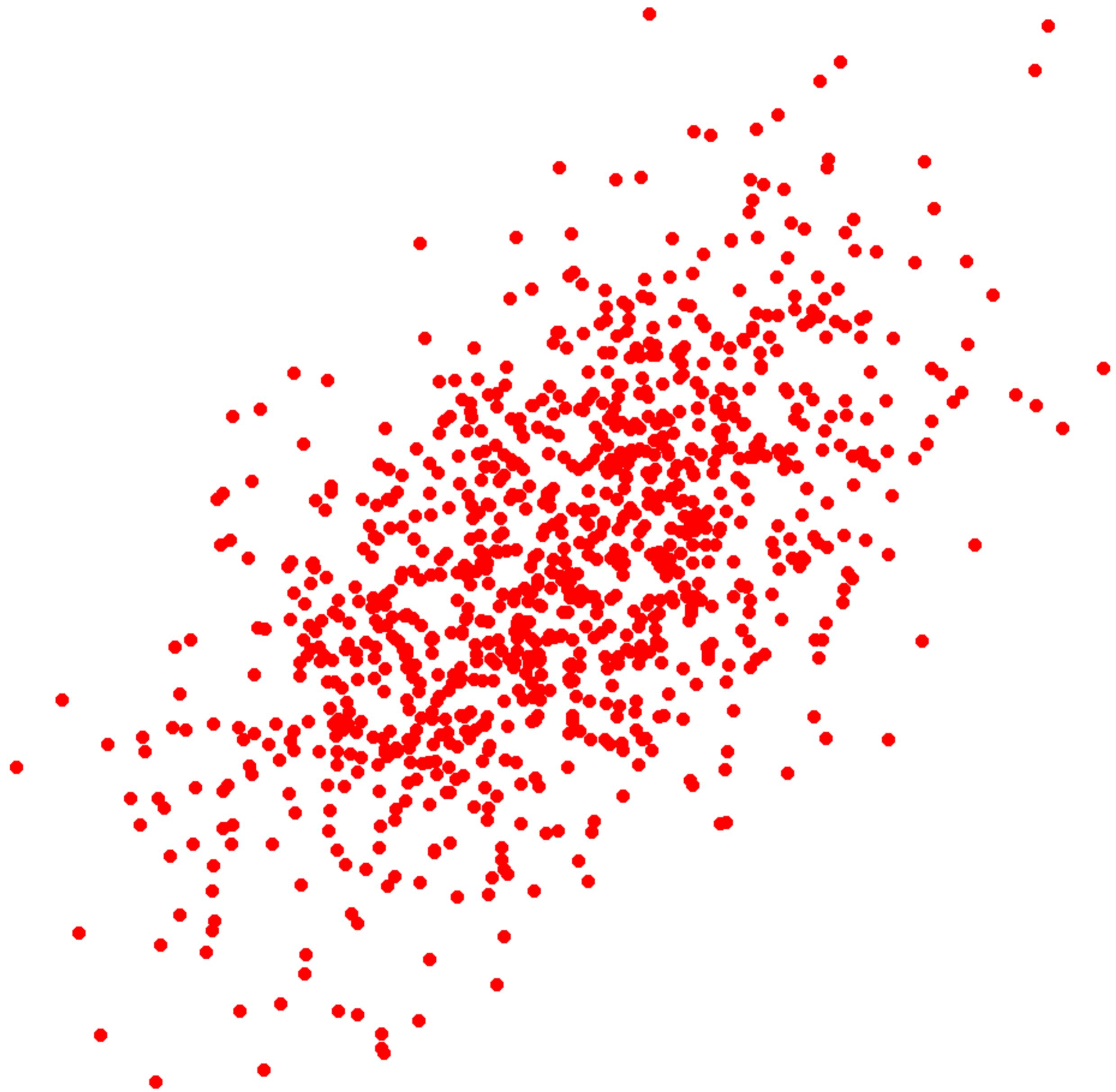
0.56



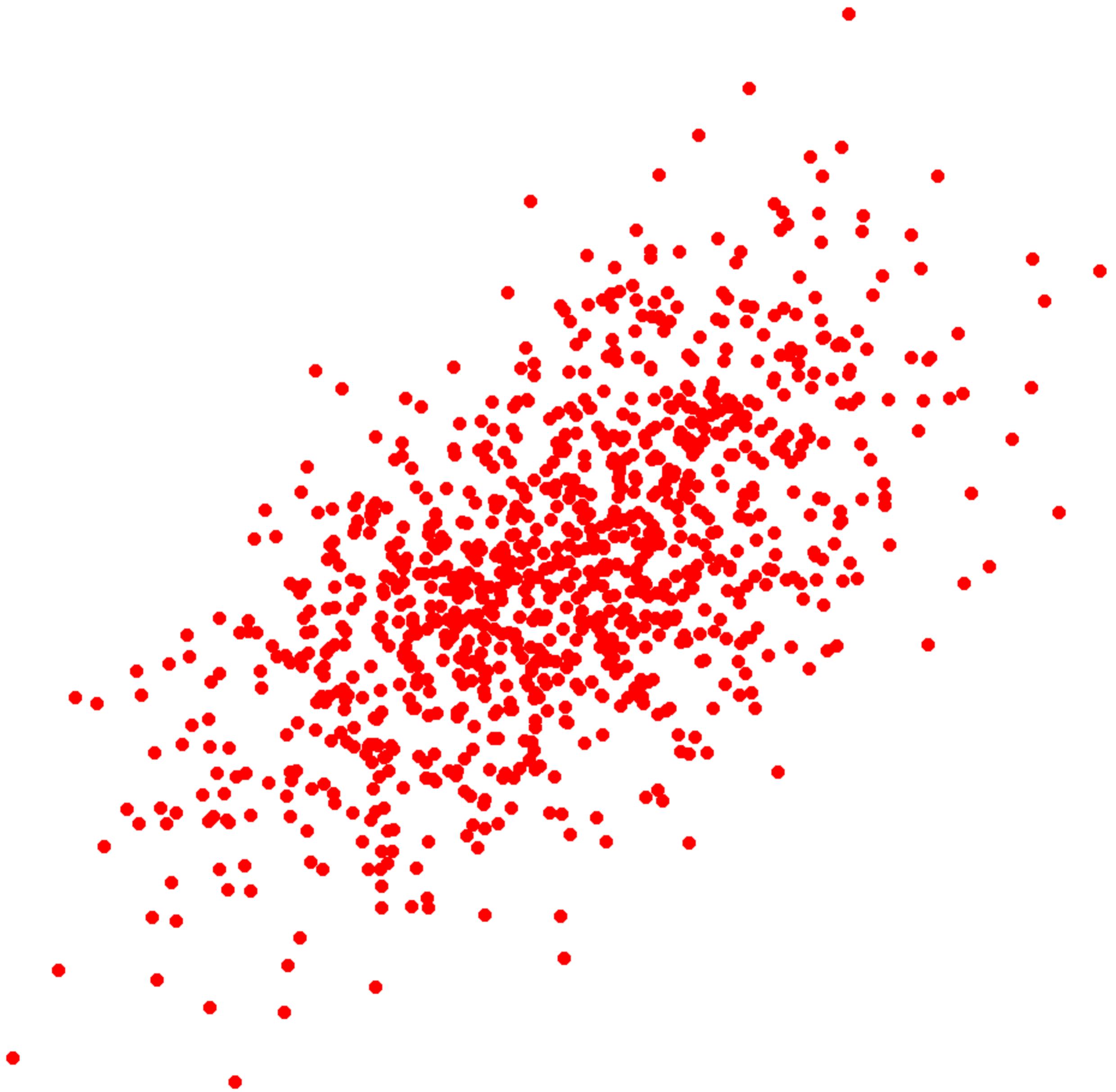
0.57



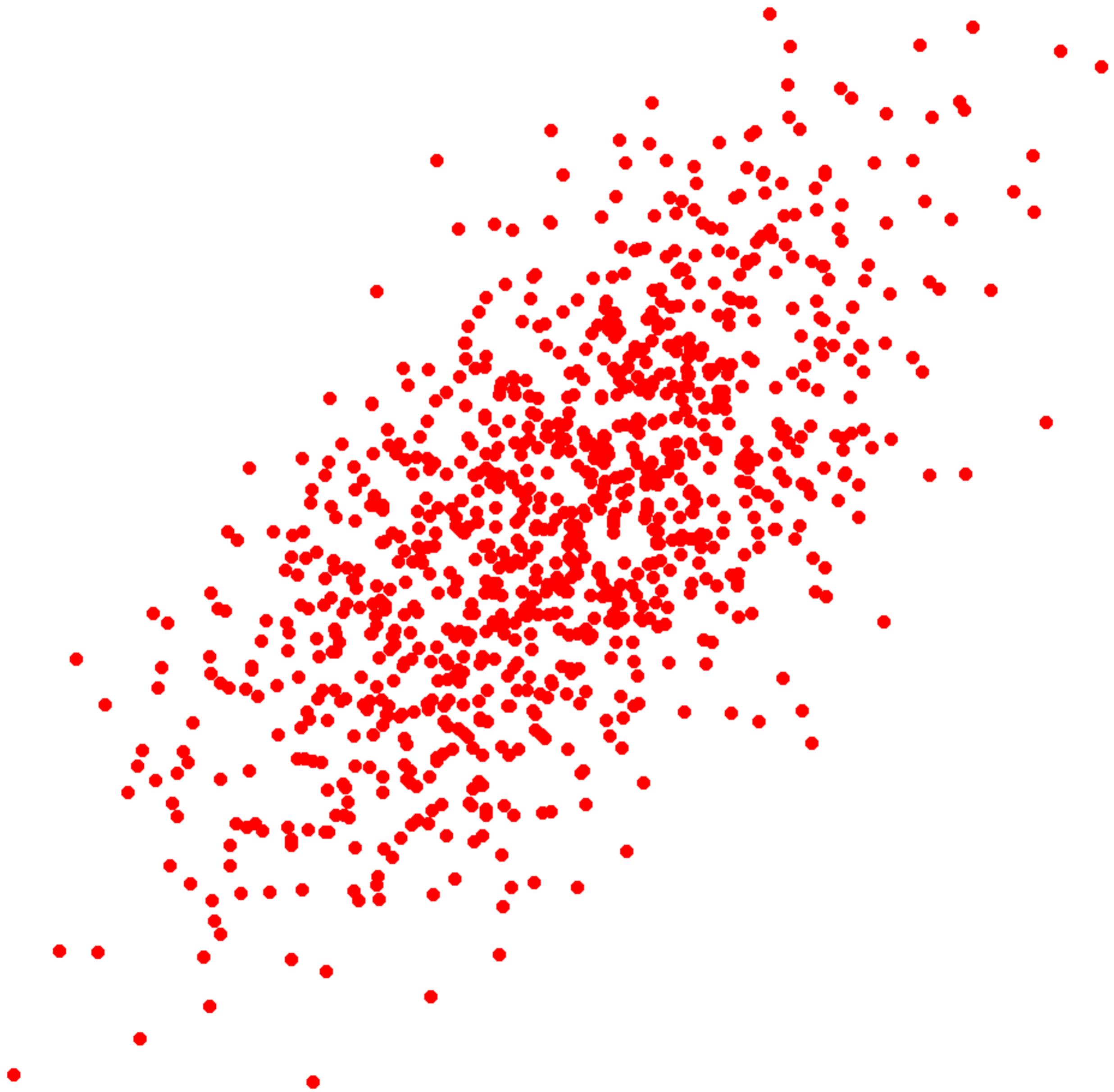
0.57



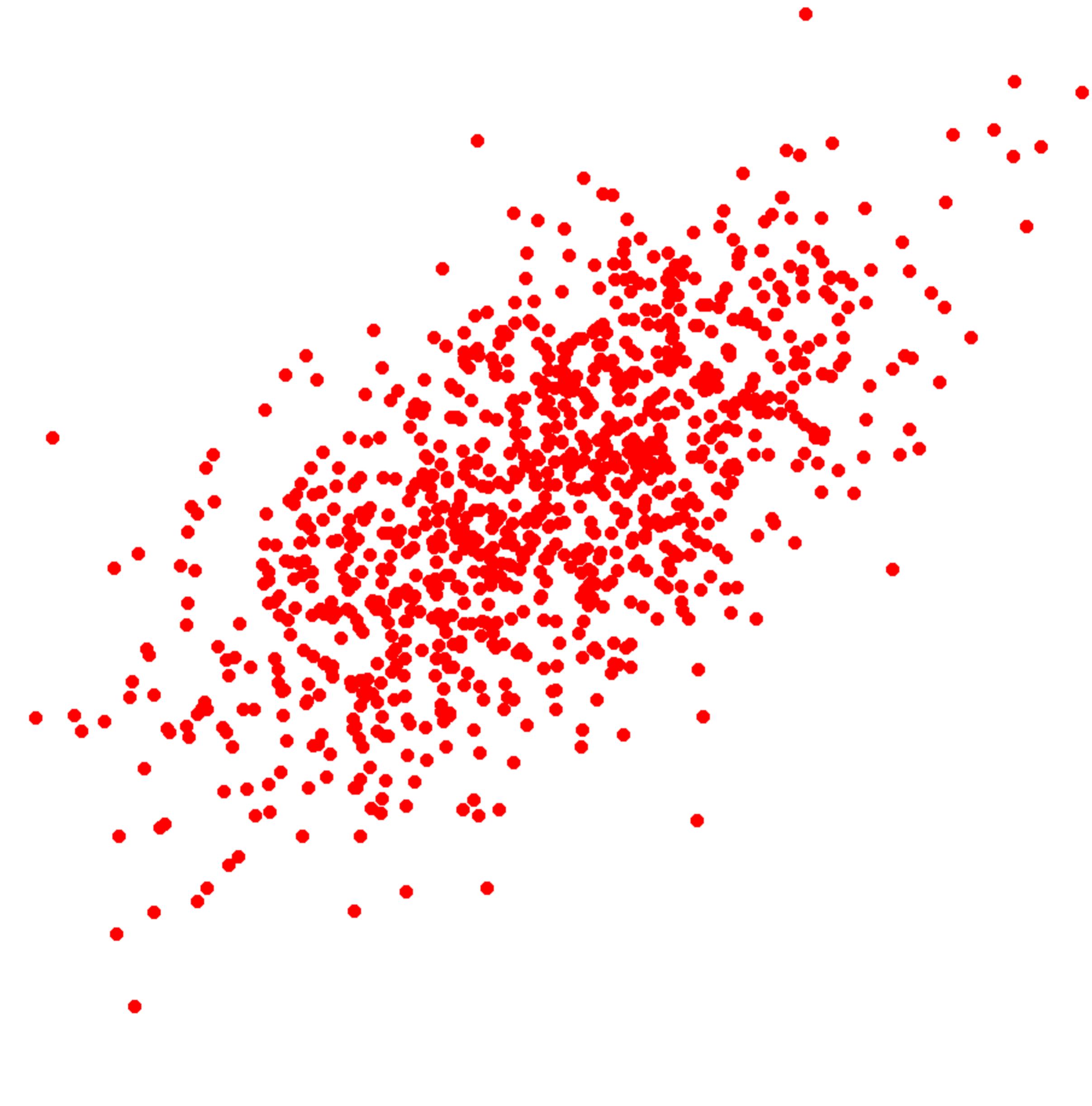
0.59



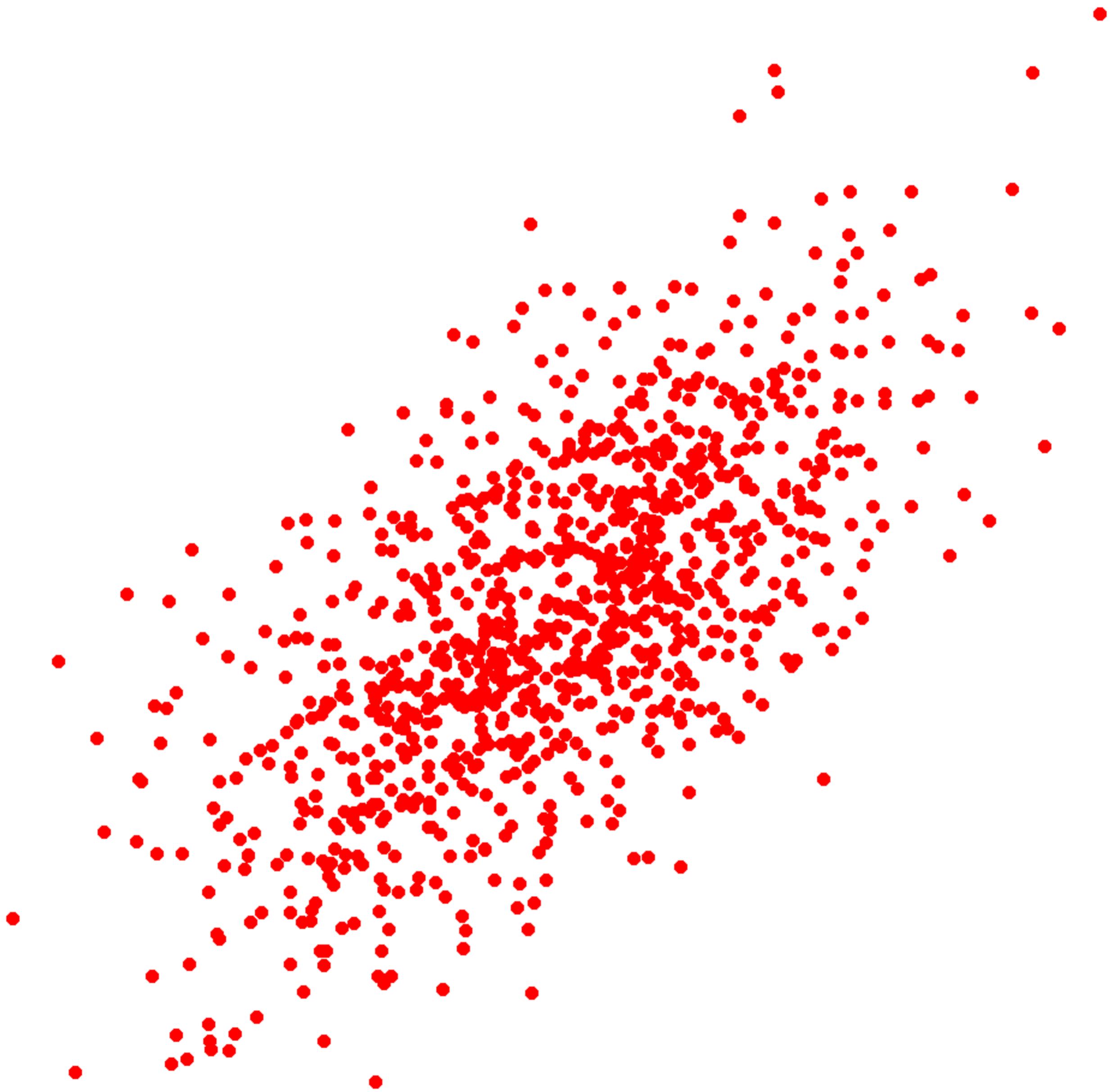
0.64



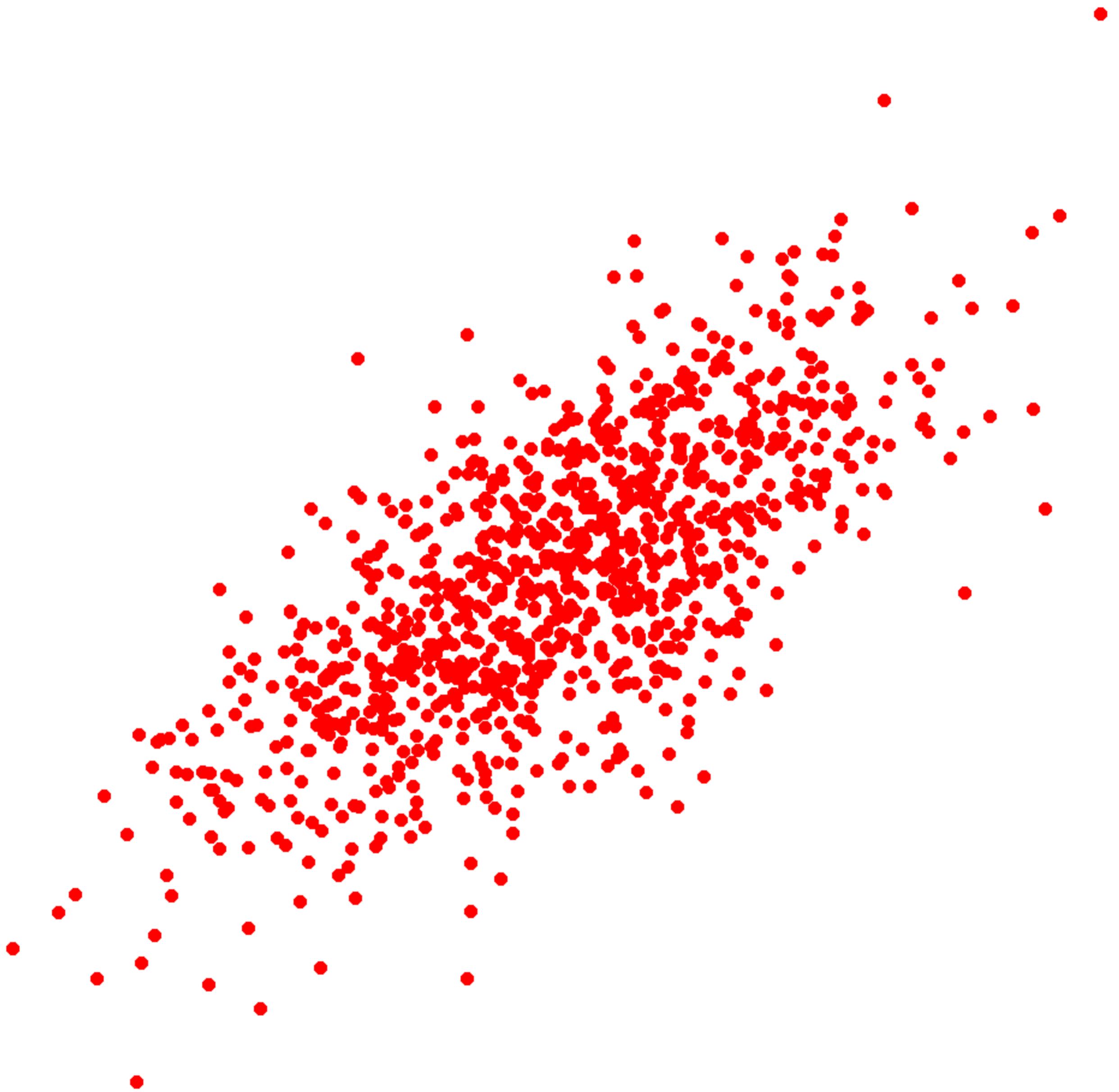
0.64



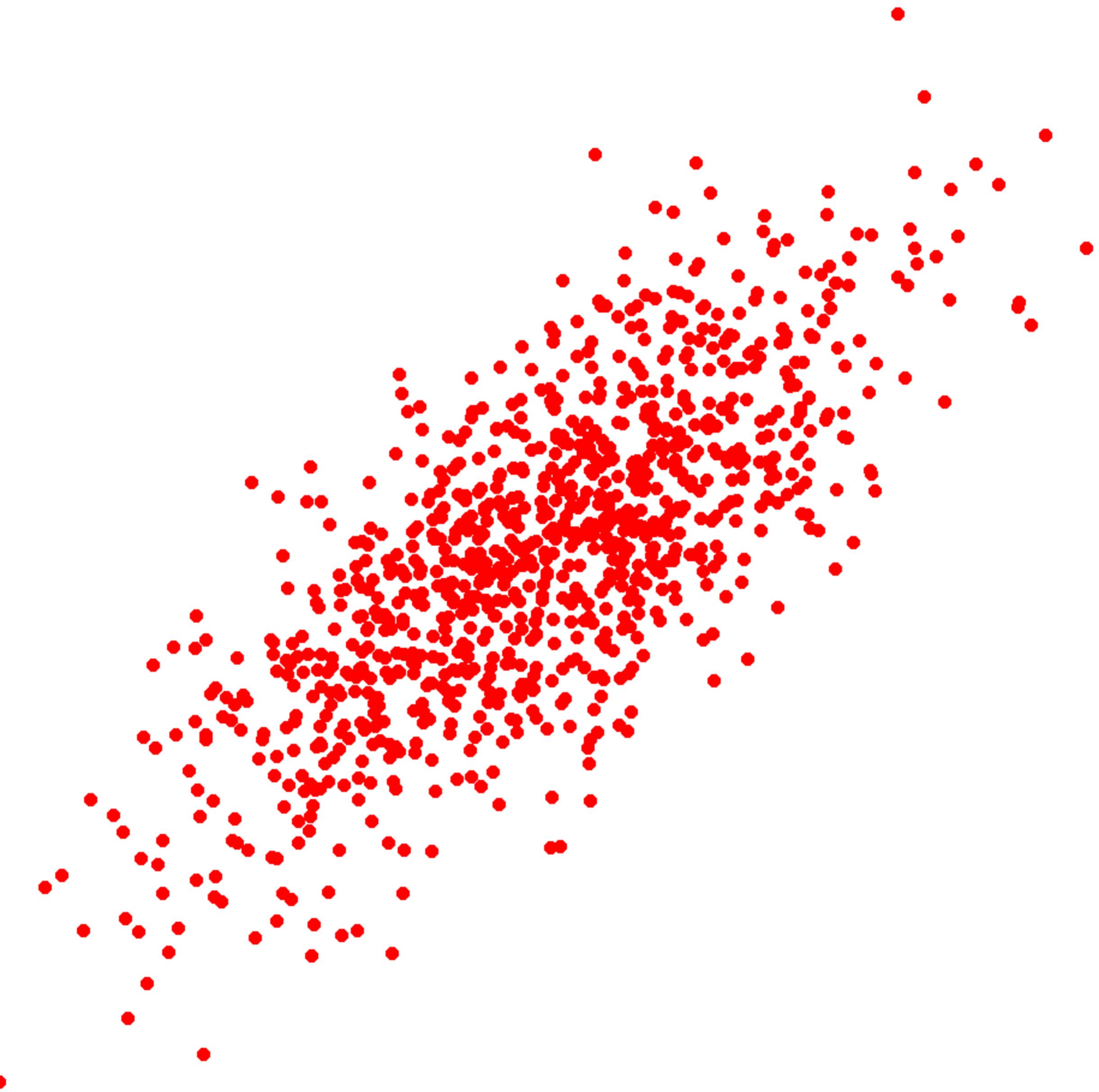
0.66



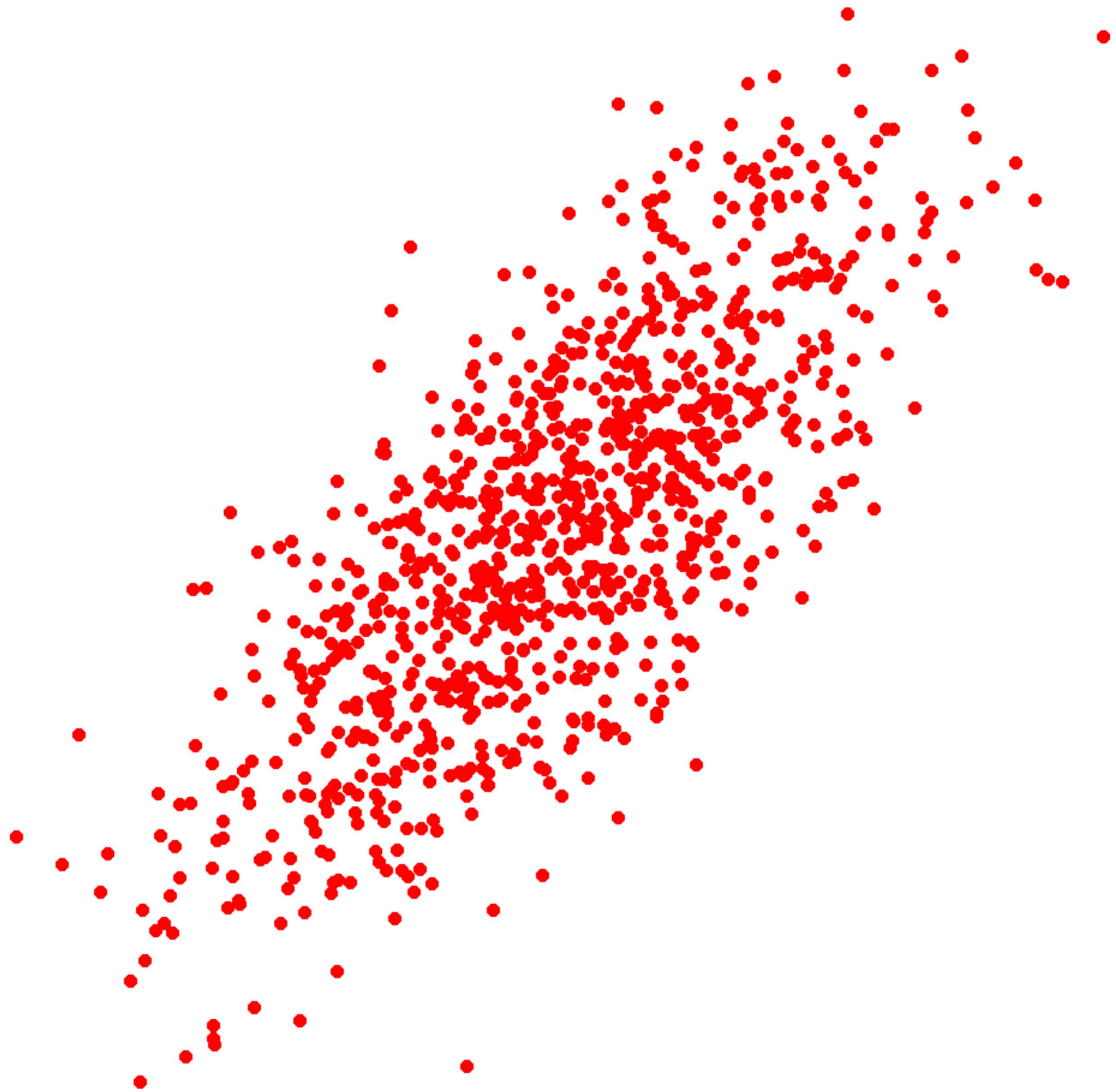
0.71



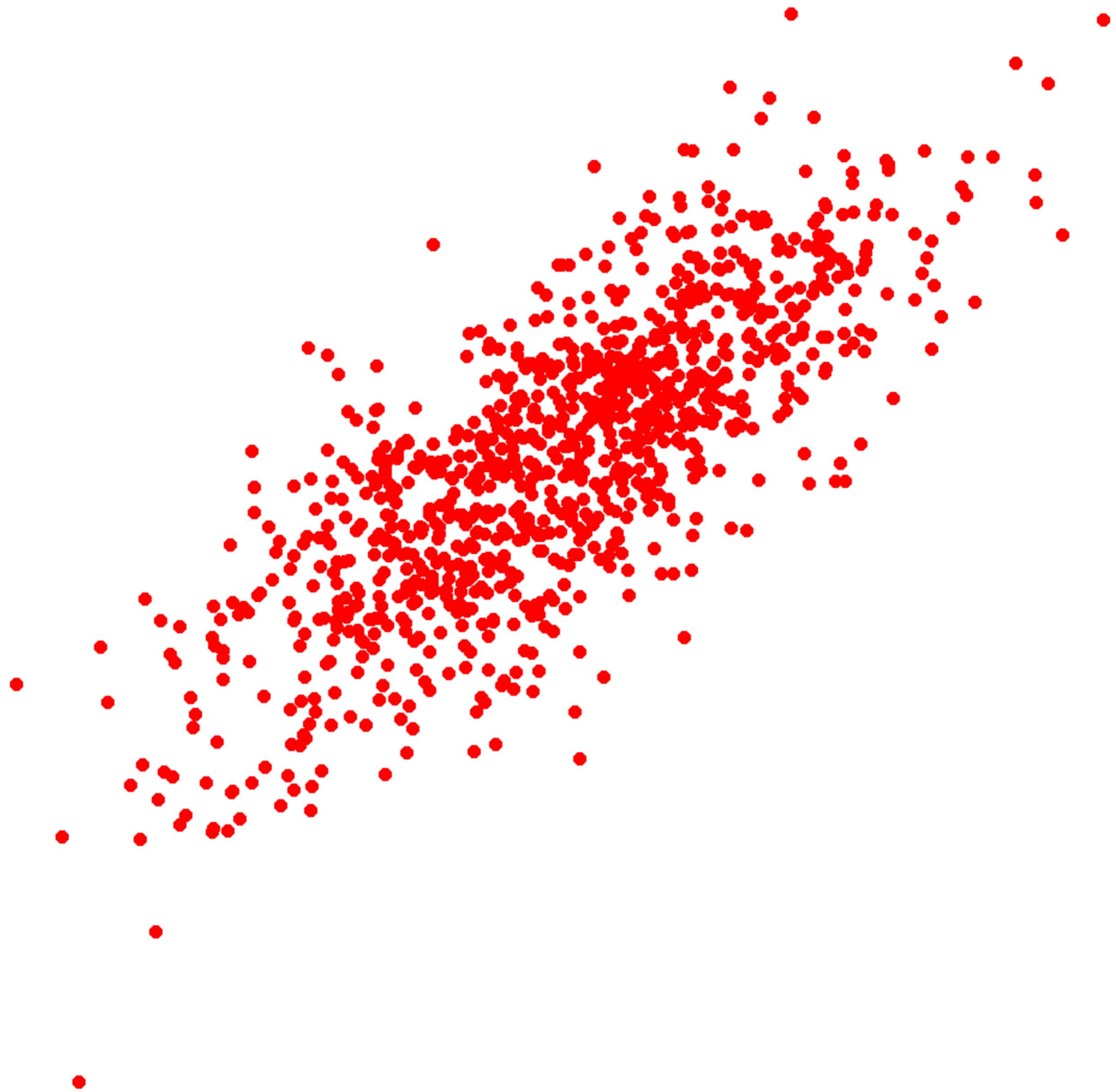
0.75



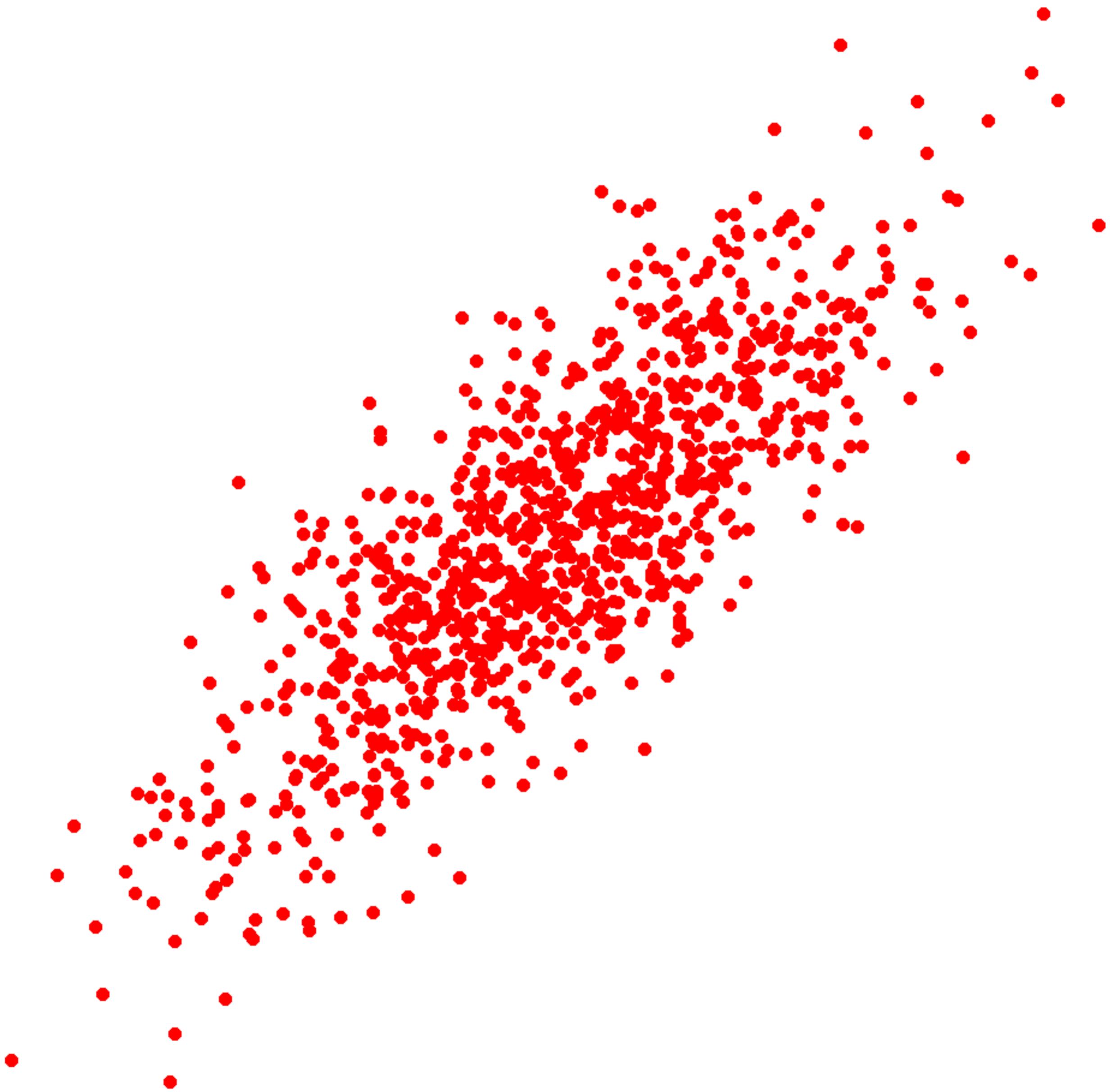
0.75



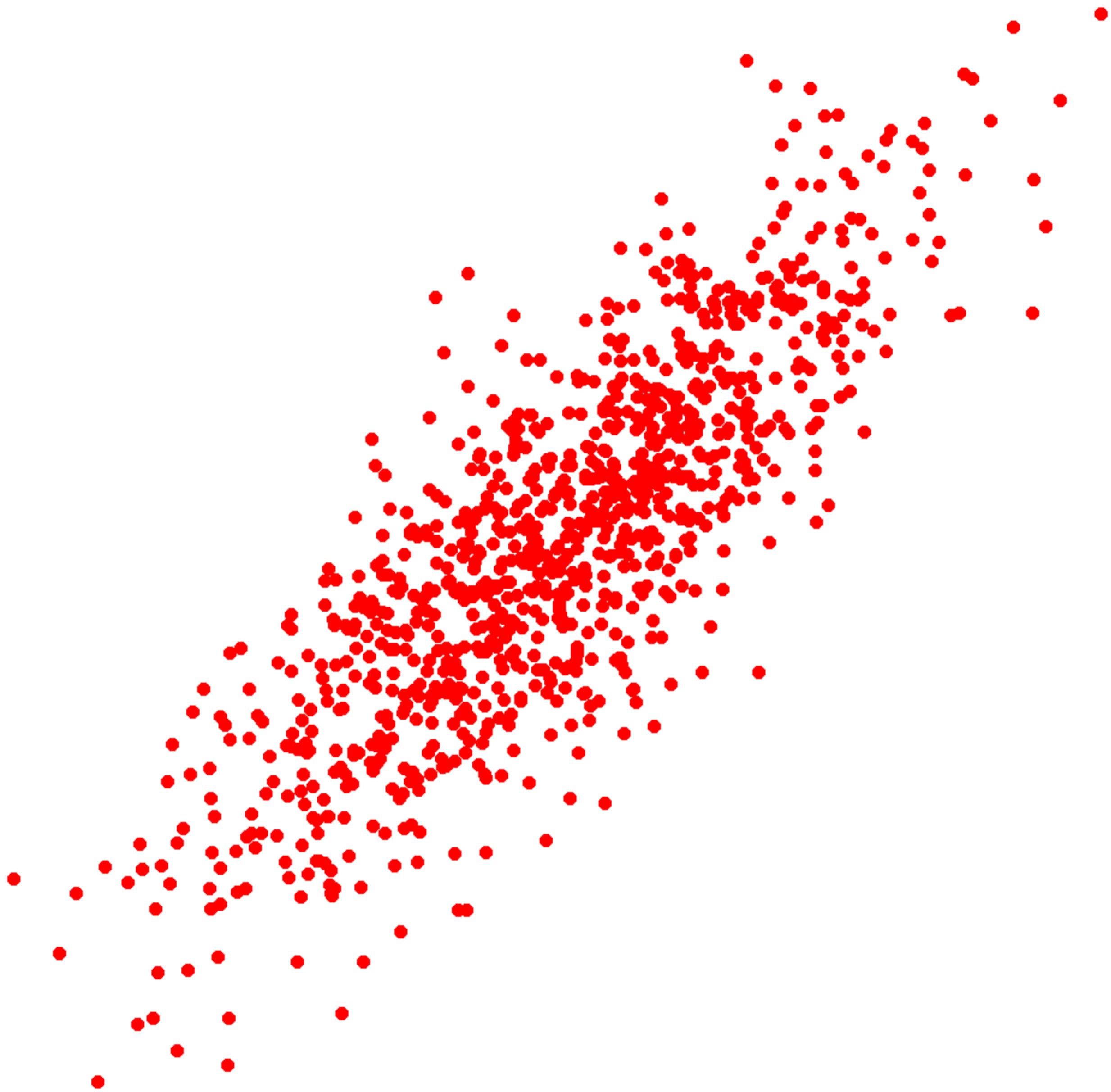
0.78



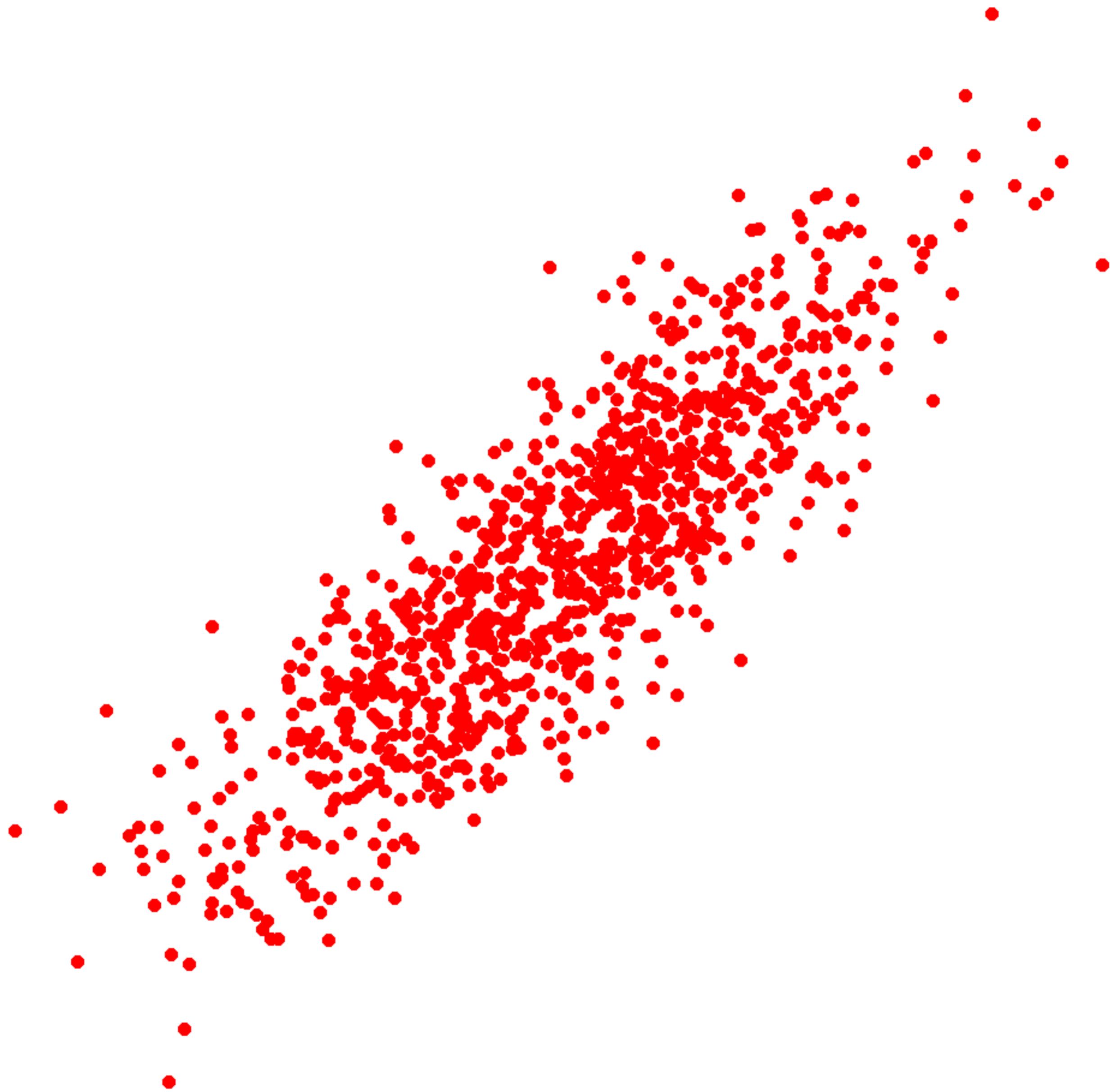
0.8



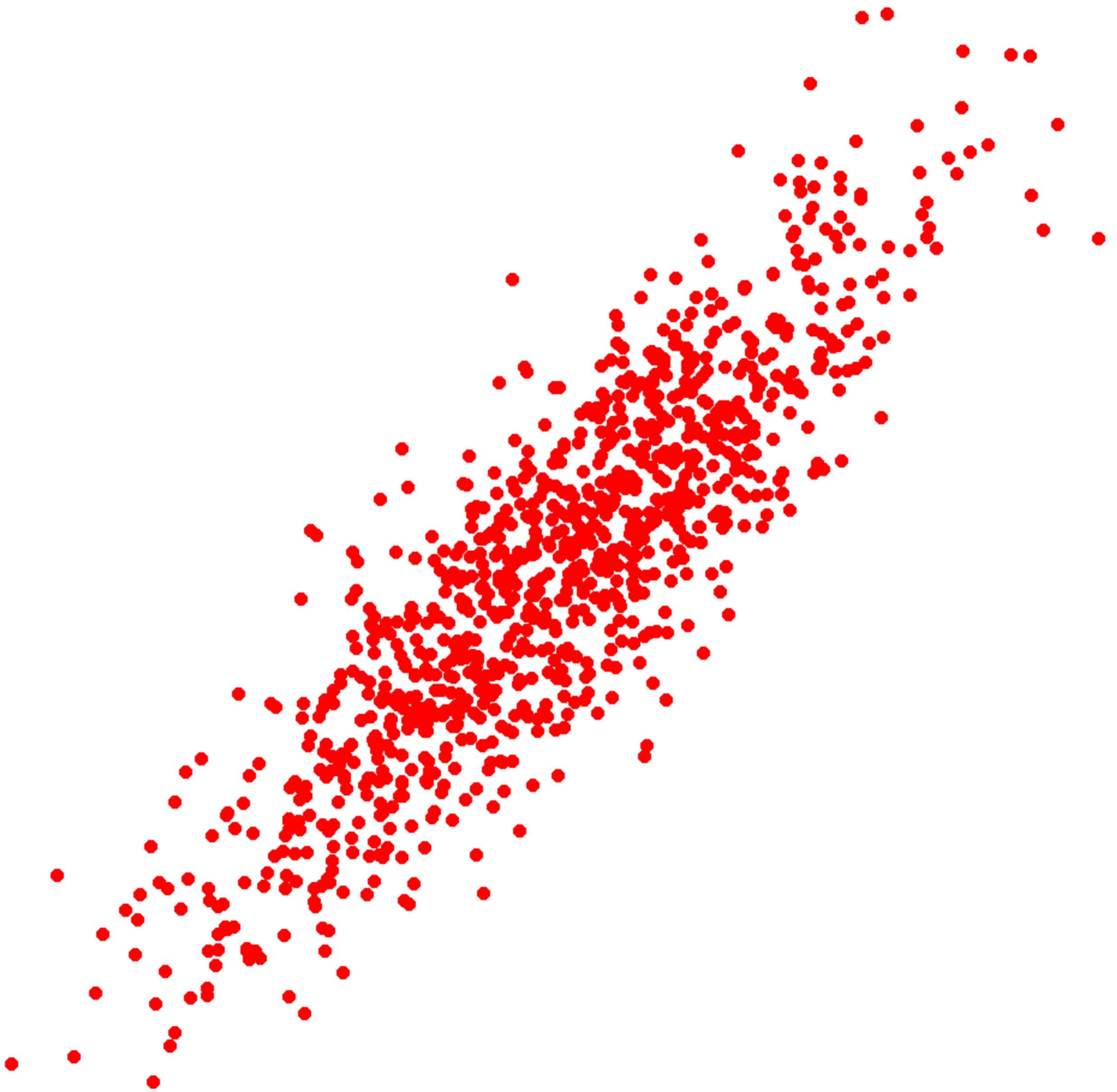
0.83



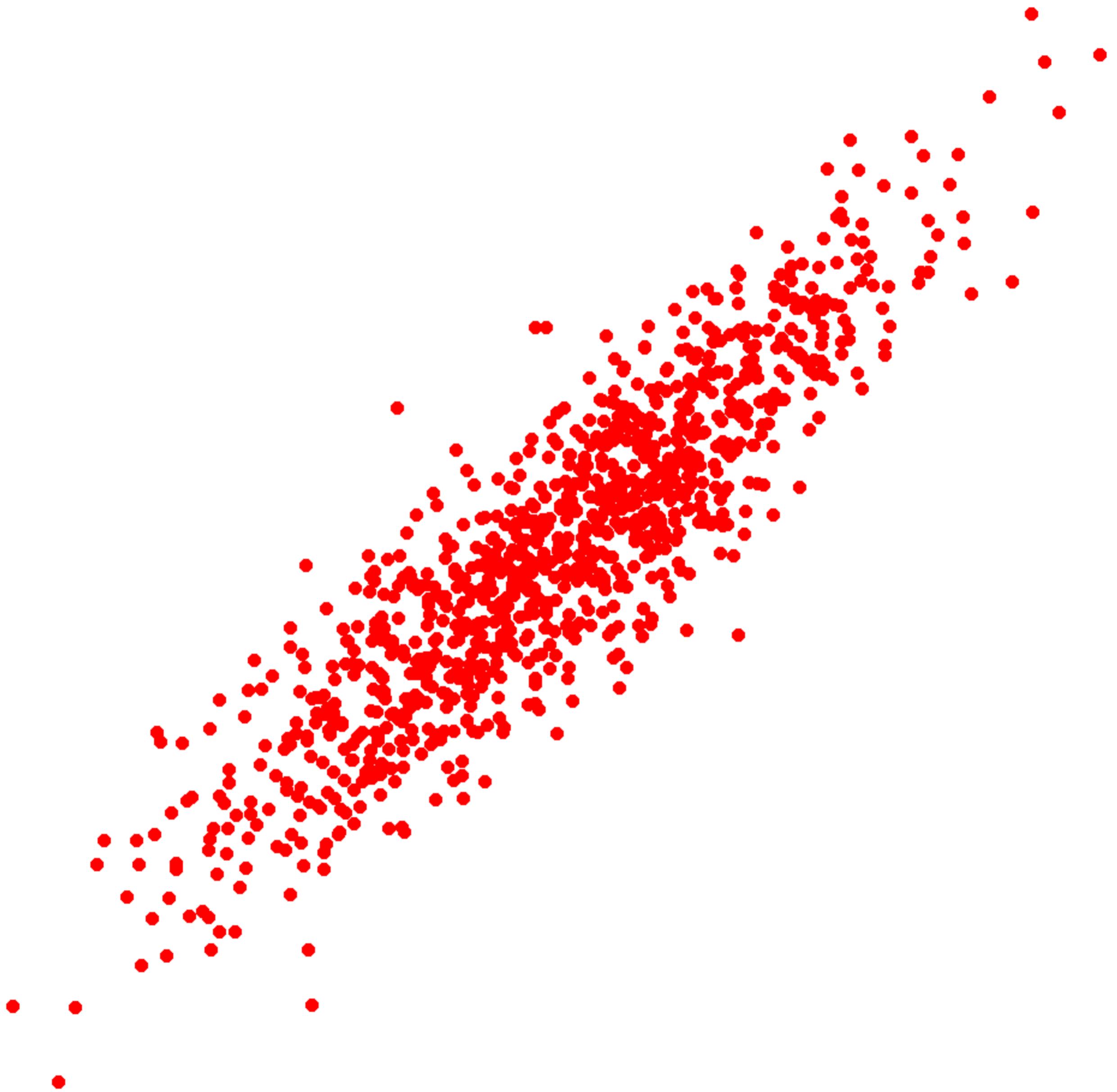
0.85



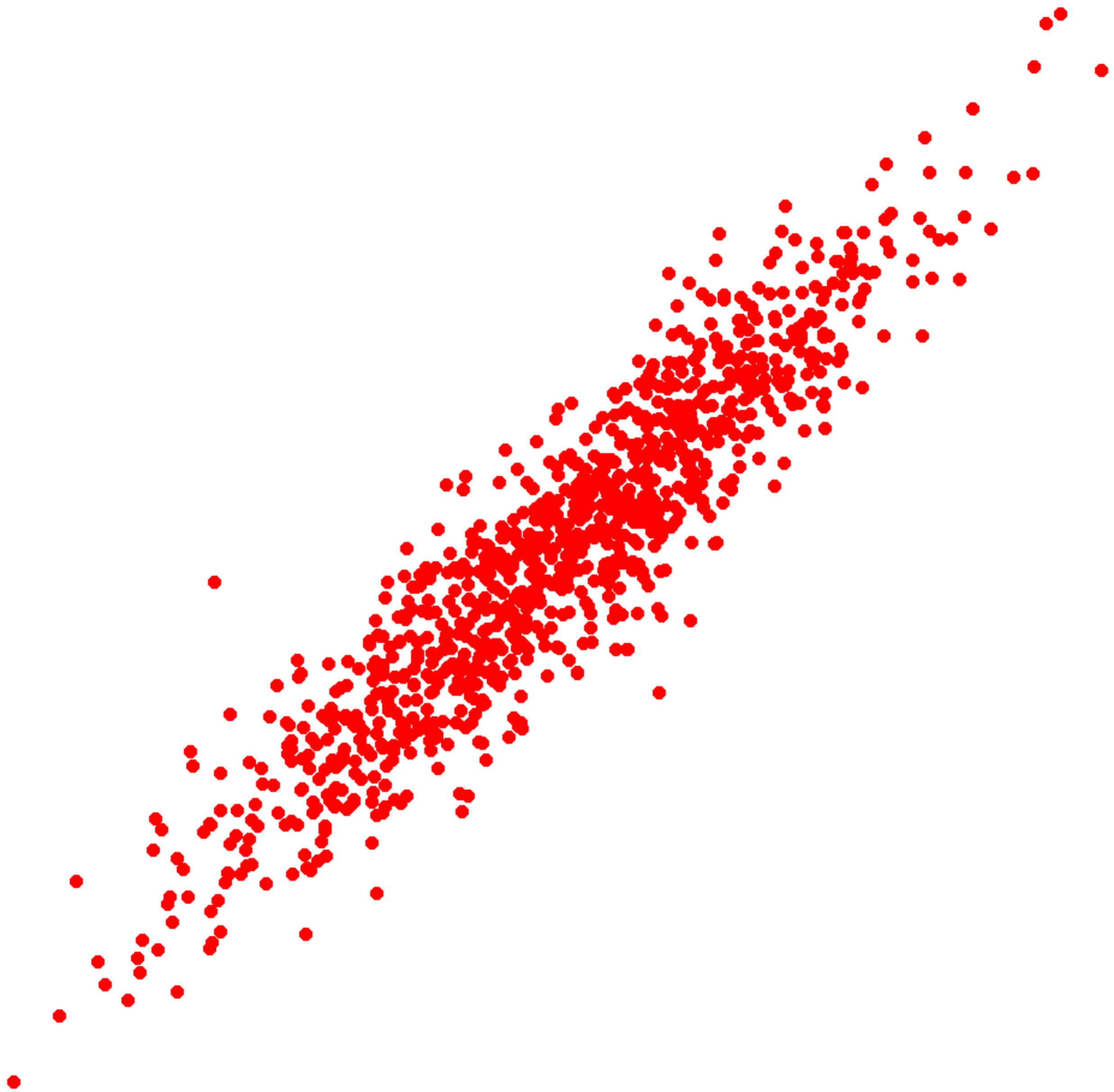
0.87



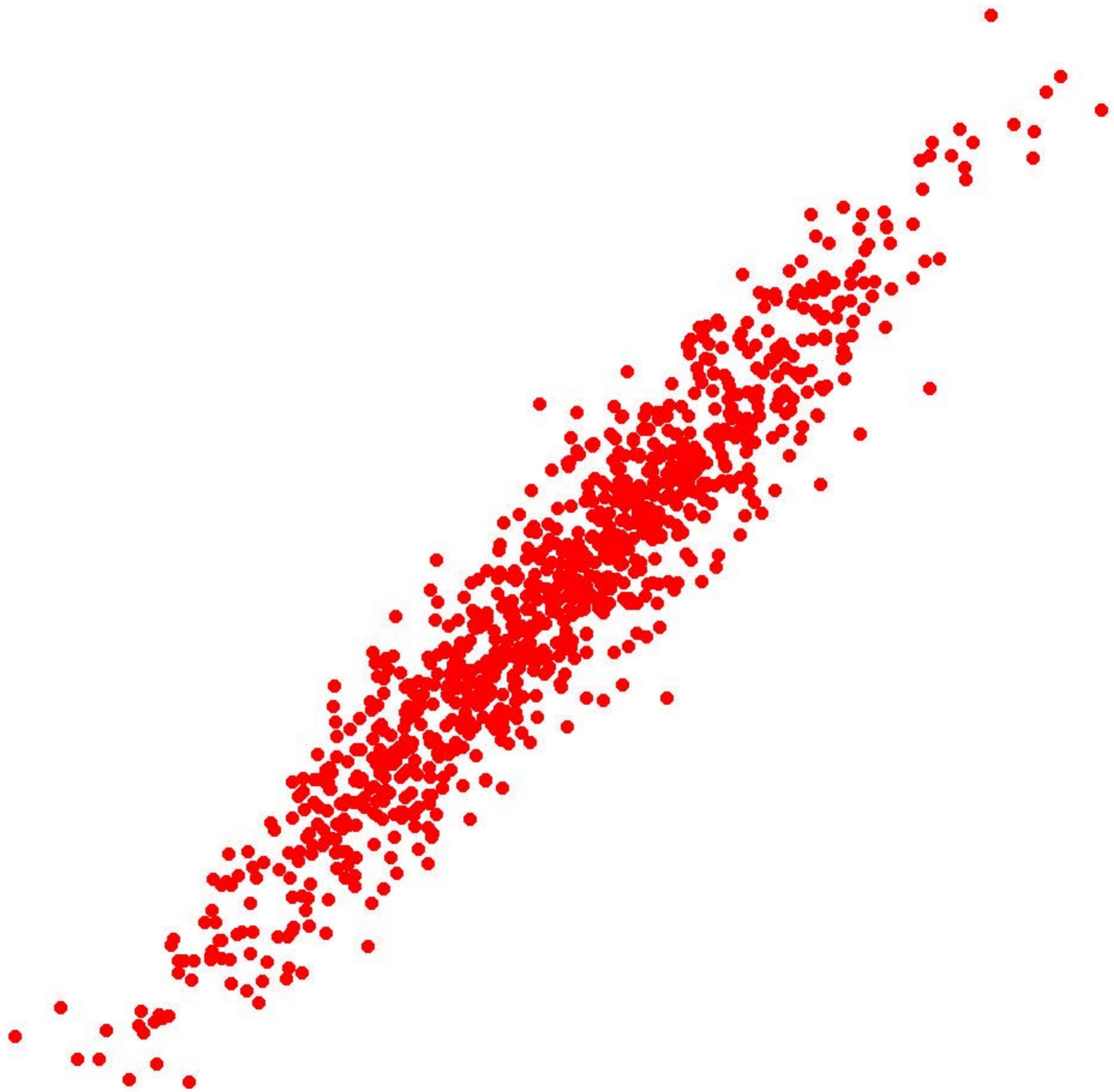
0.9



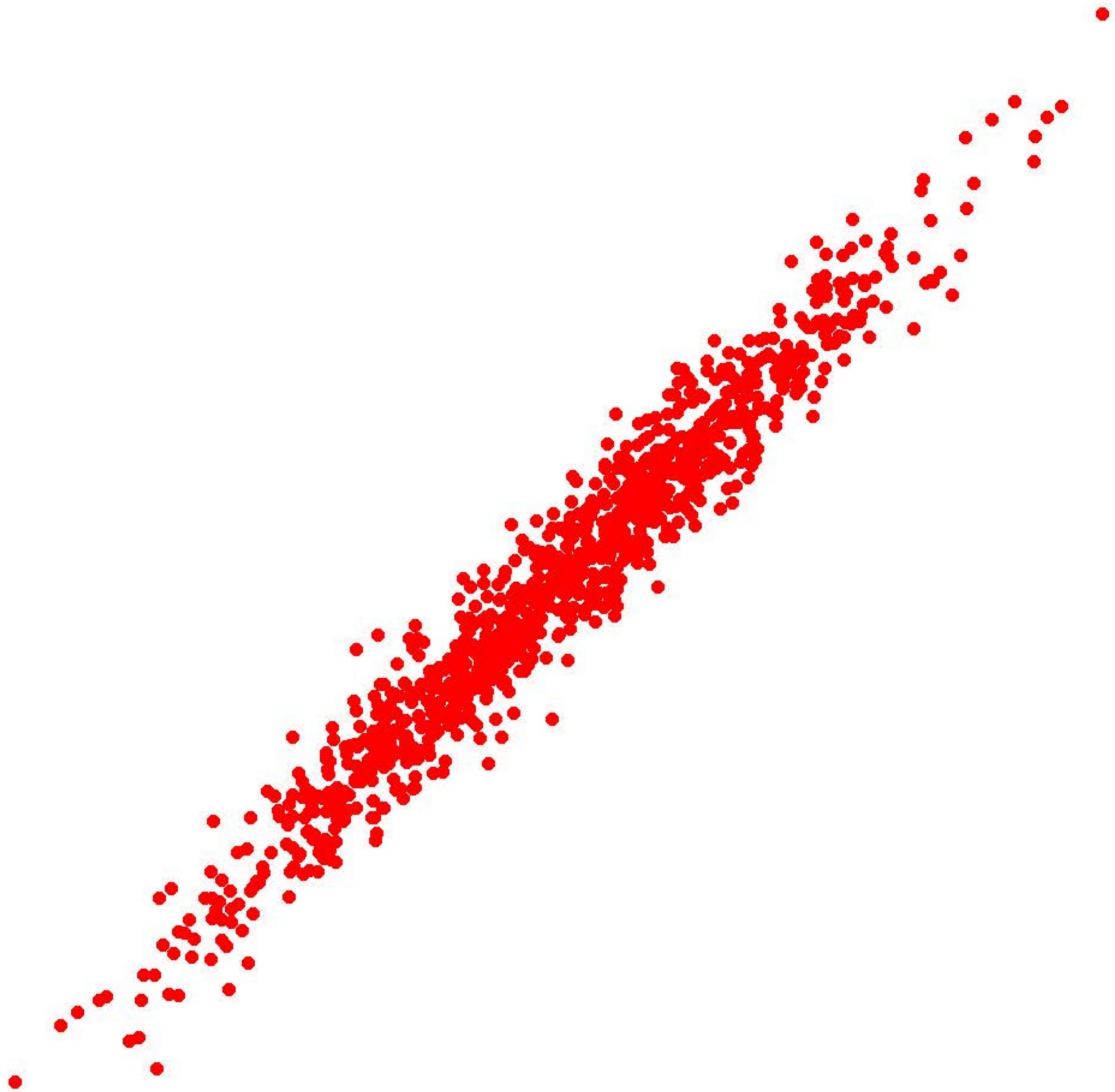
0.93



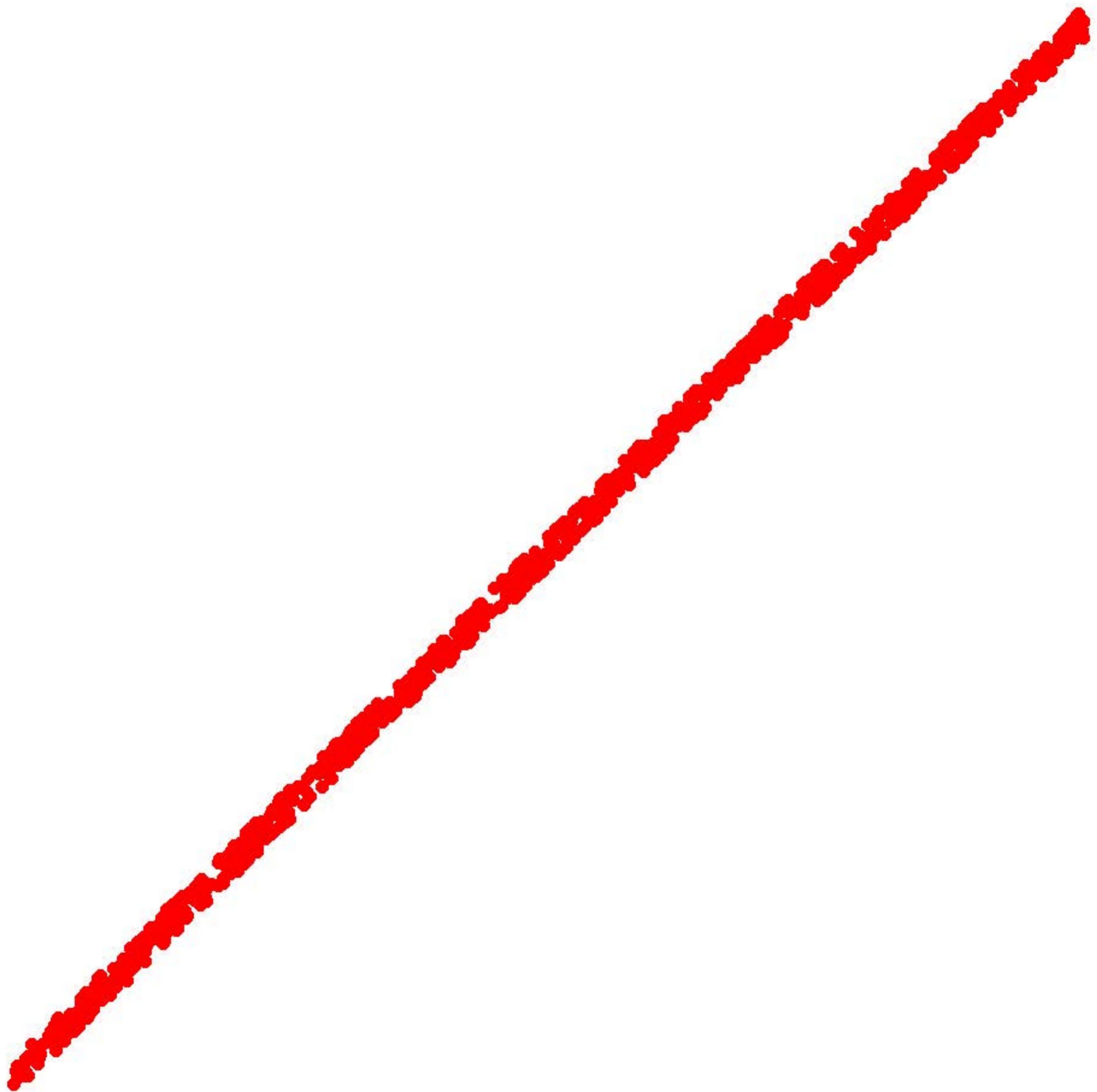
0.95



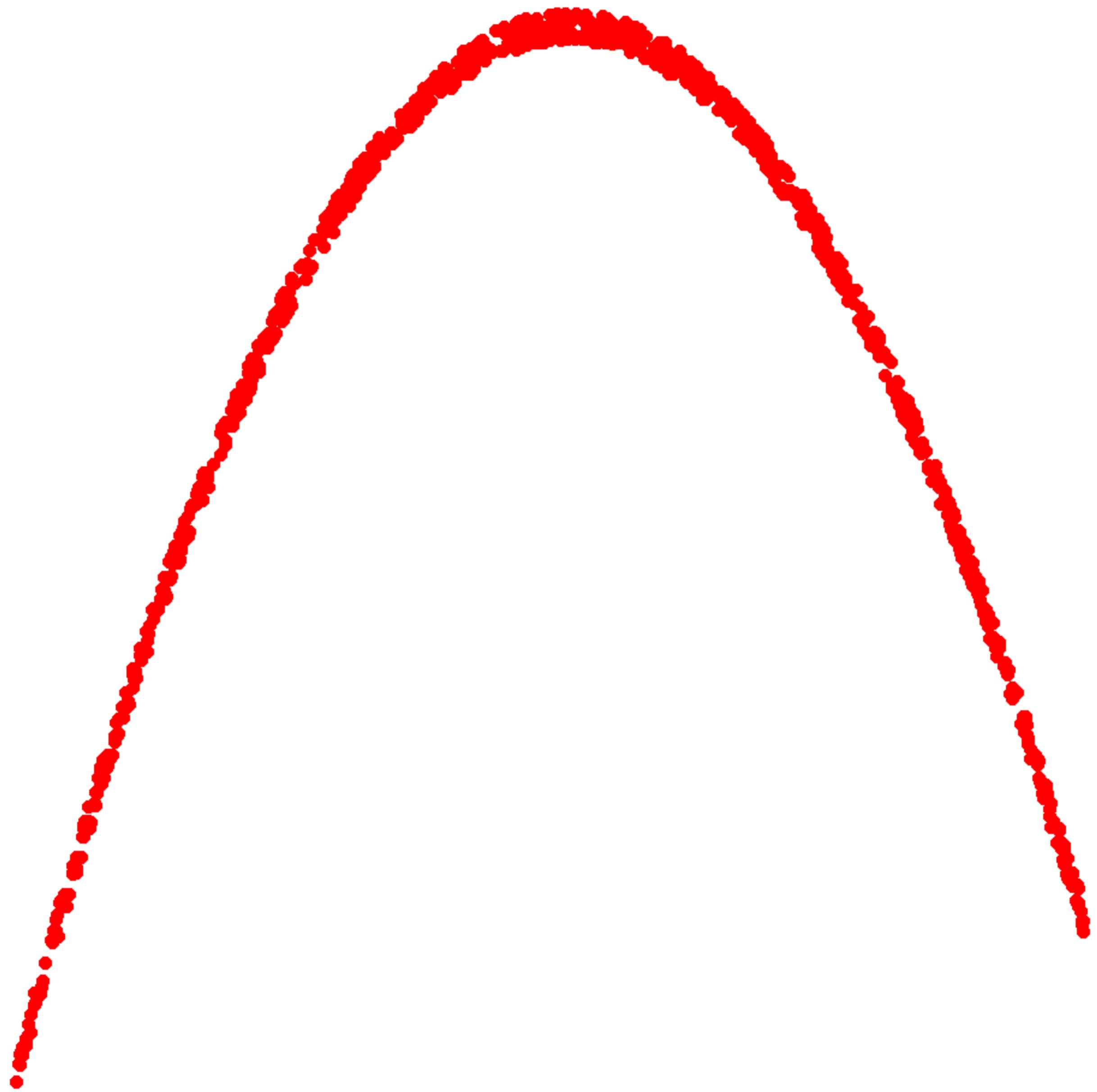
0.97



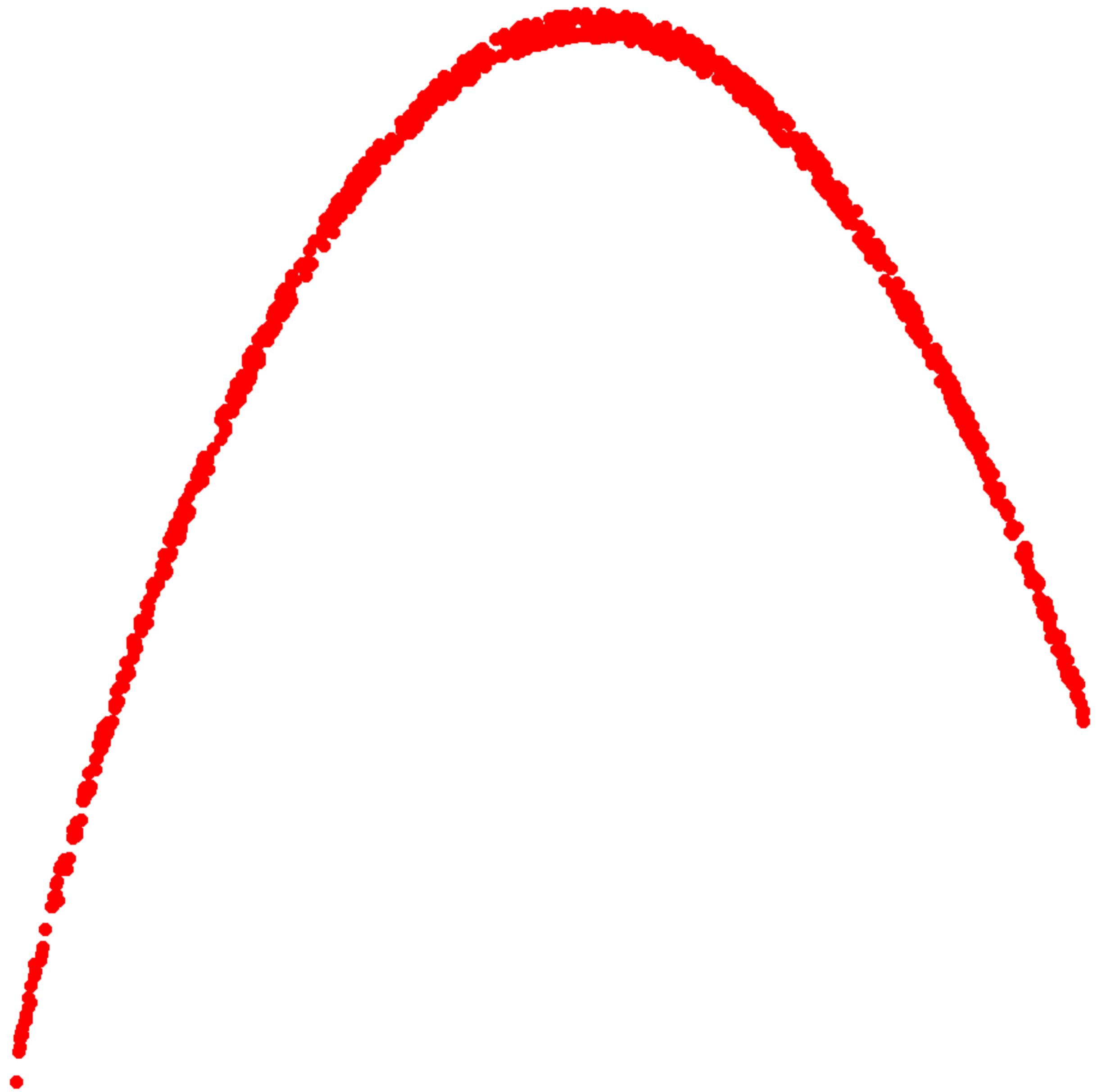
0.99958



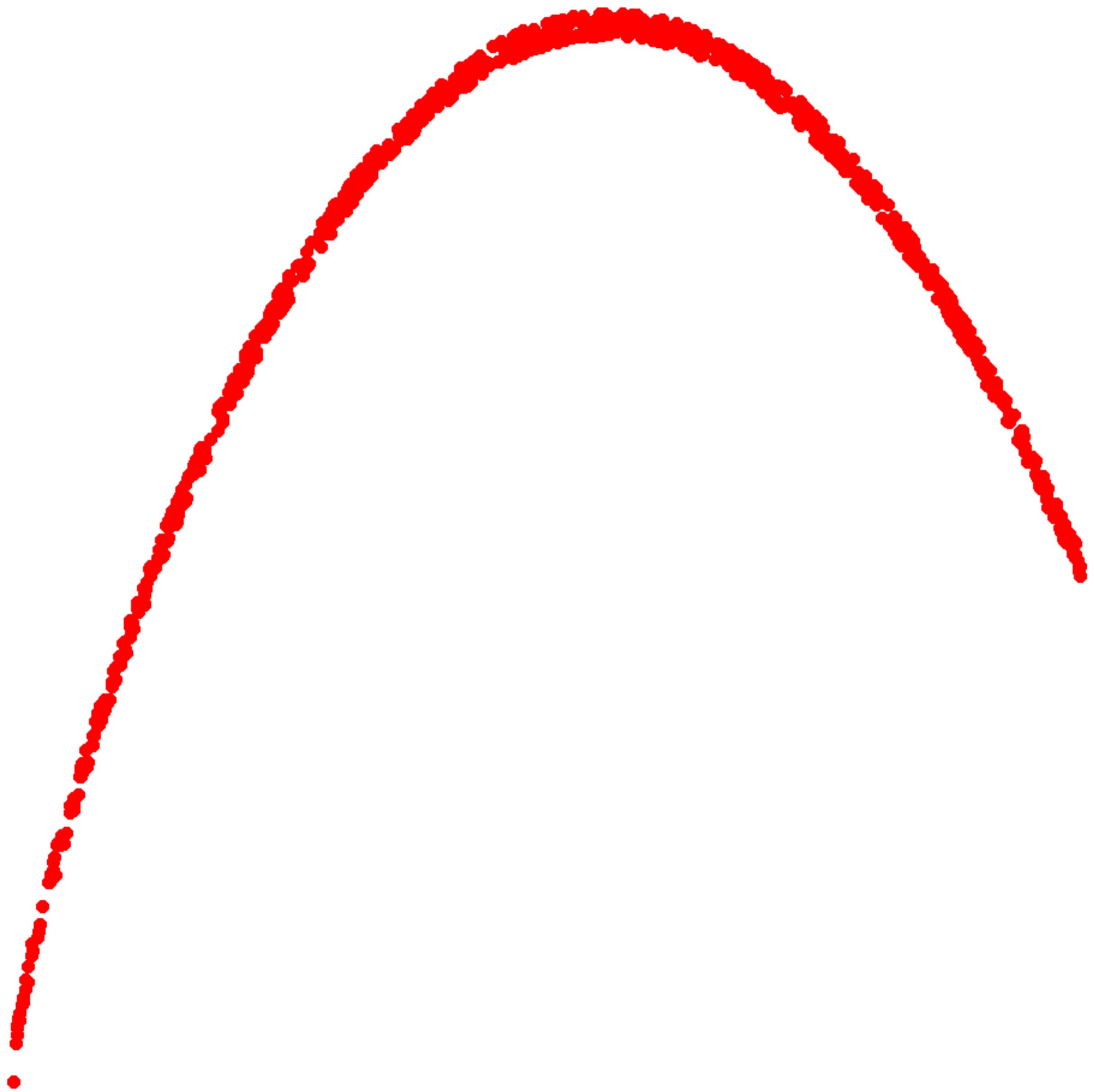
0.06161



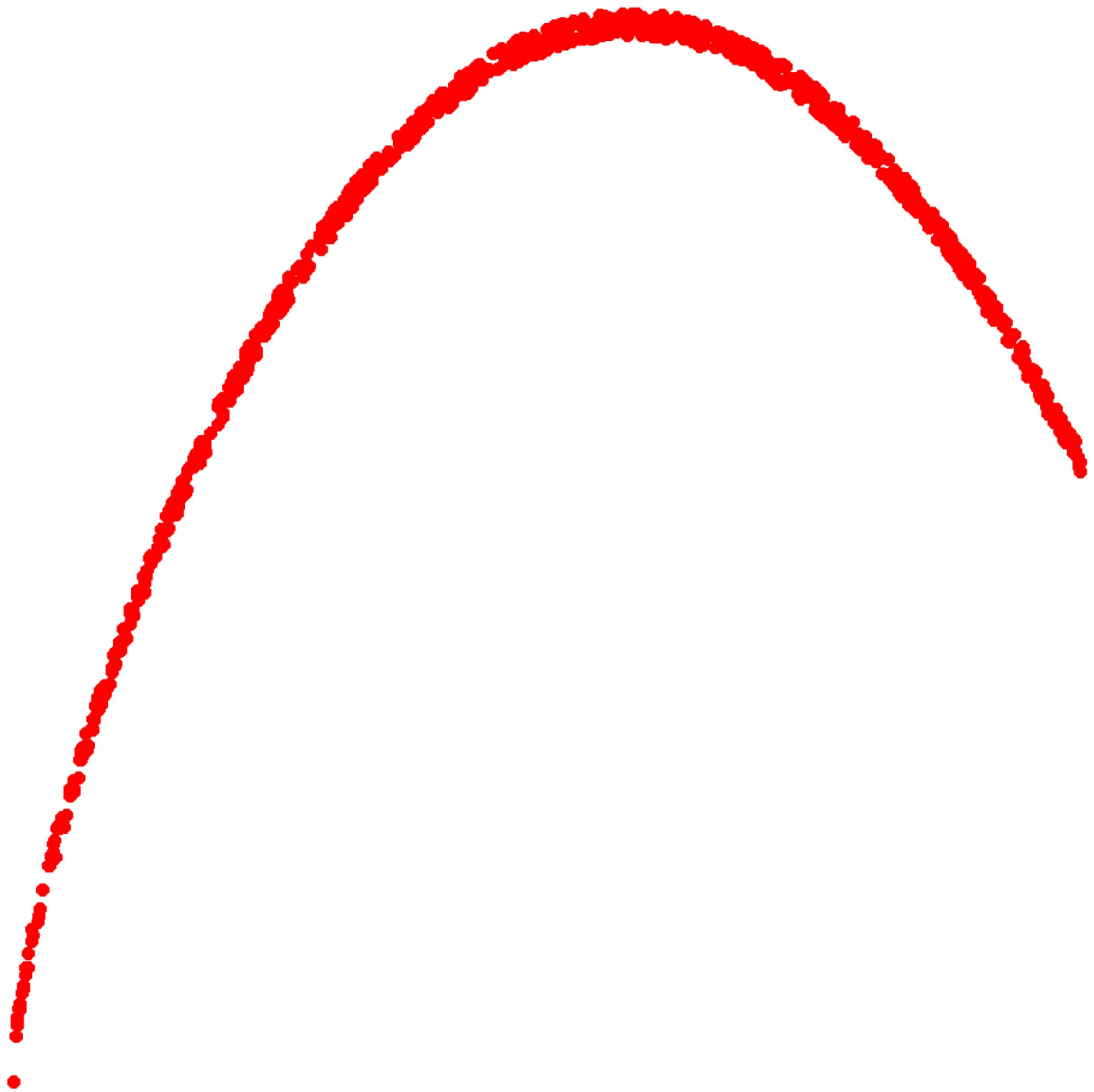
0.24377



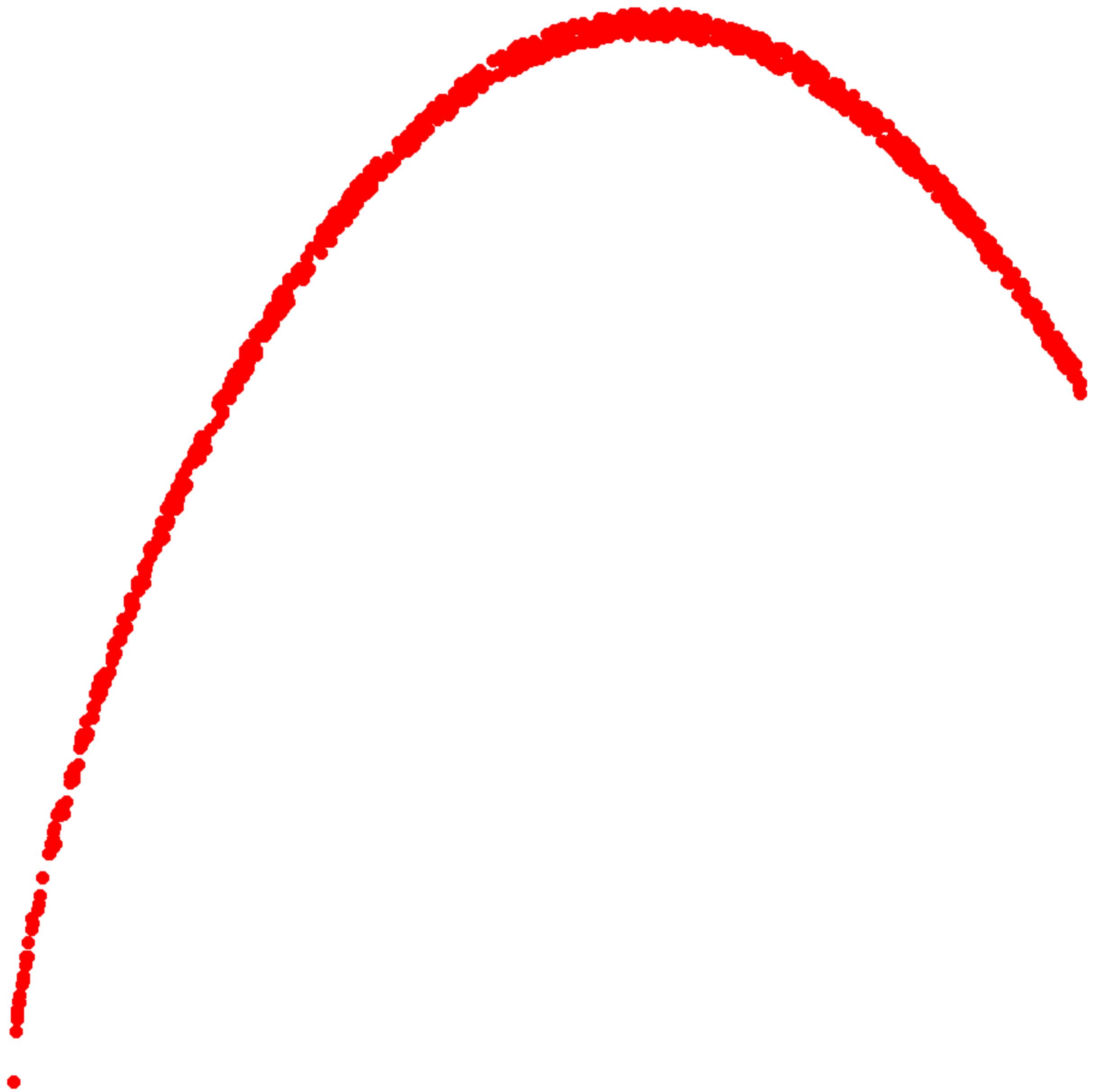
0.38846



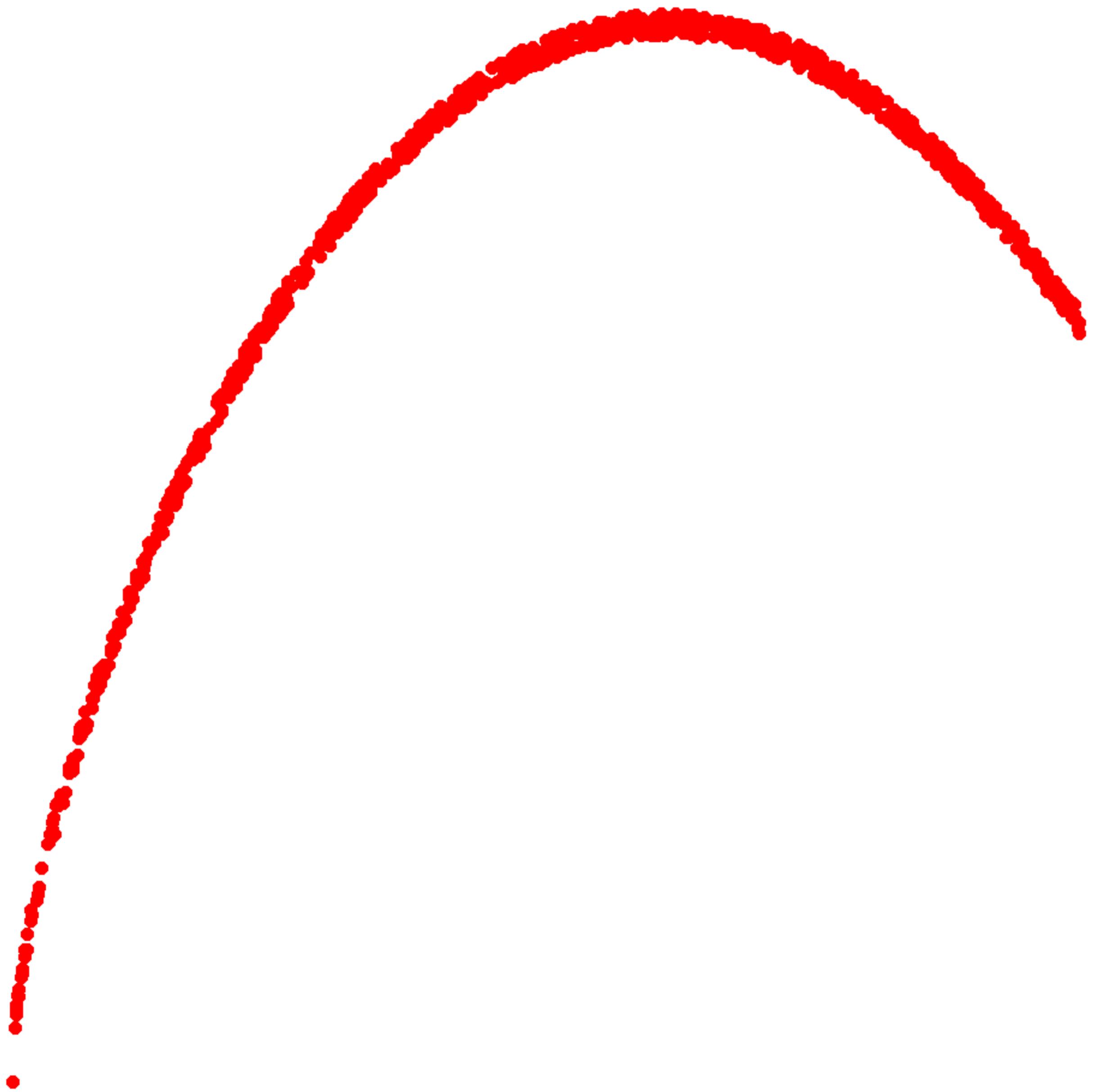
0.4984



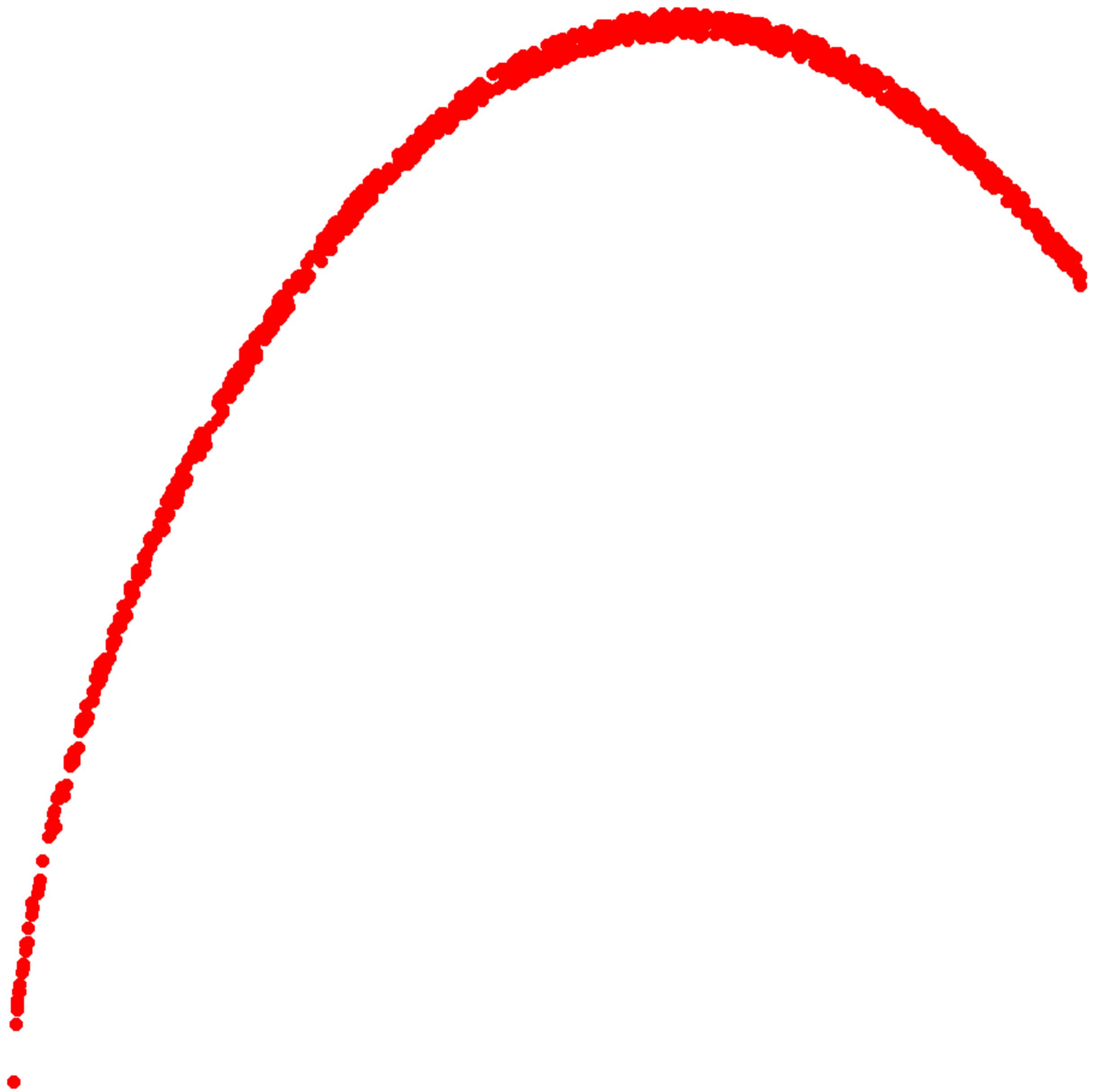
0.58082



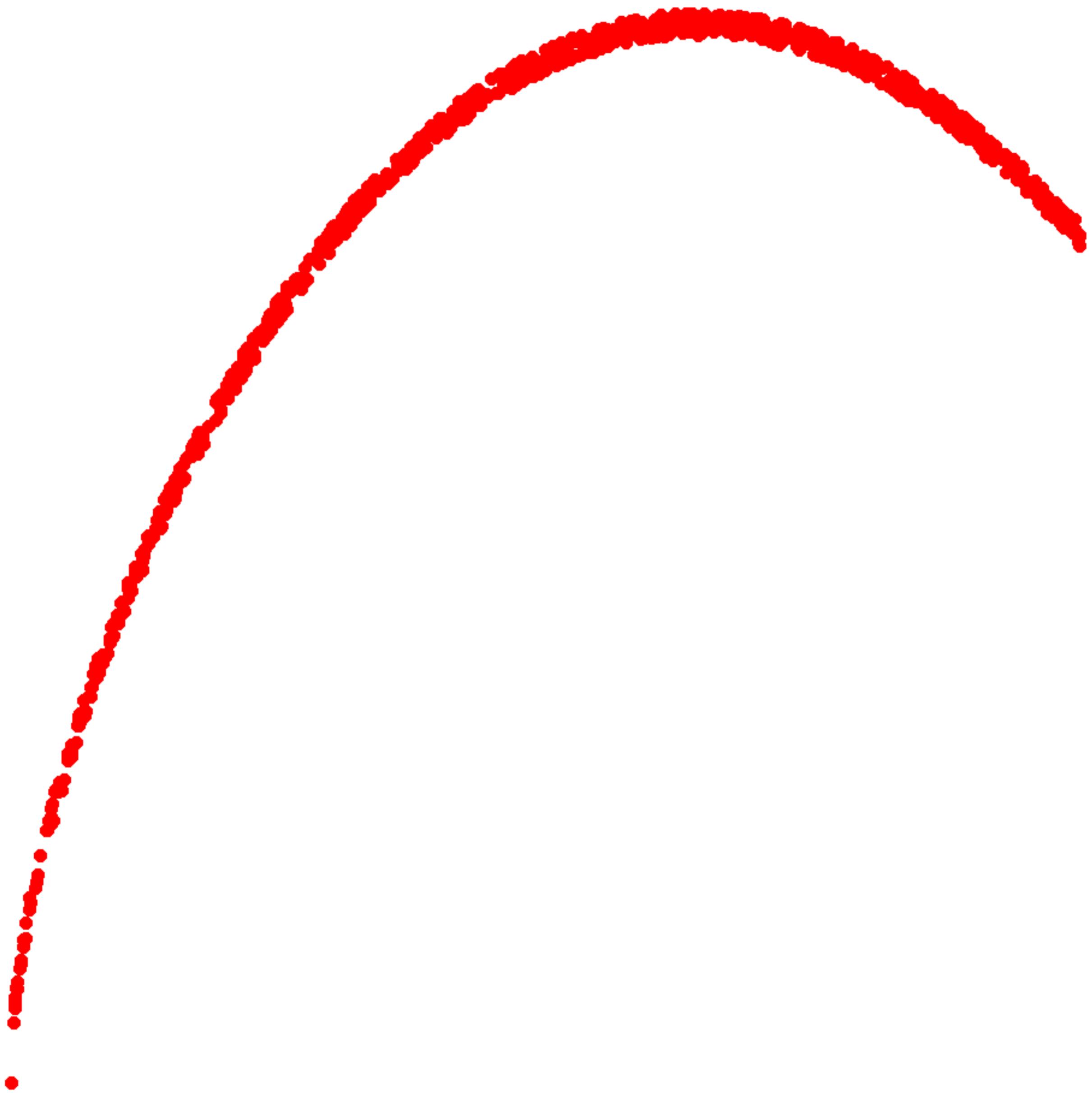
0.64287



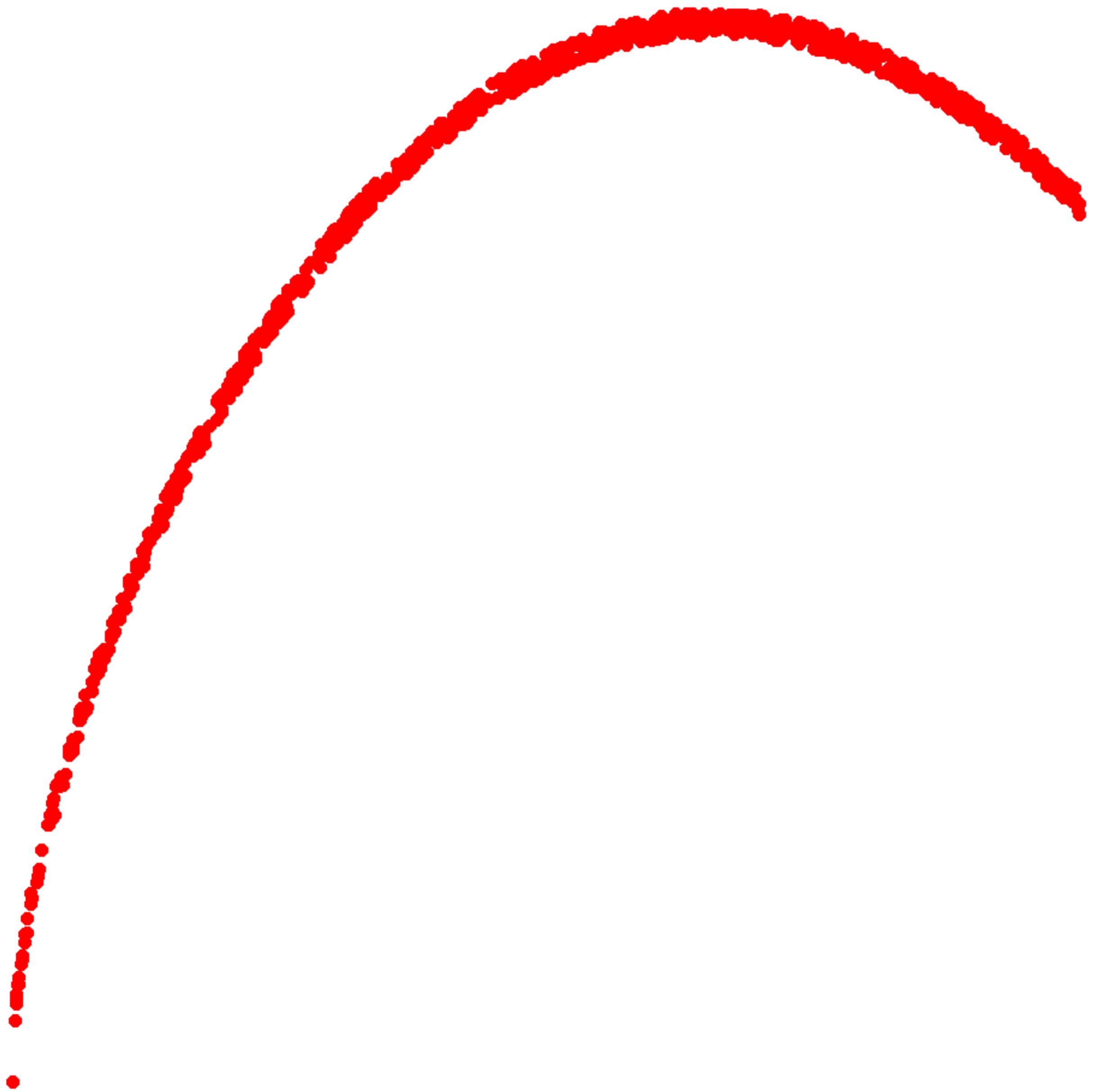
0.69019



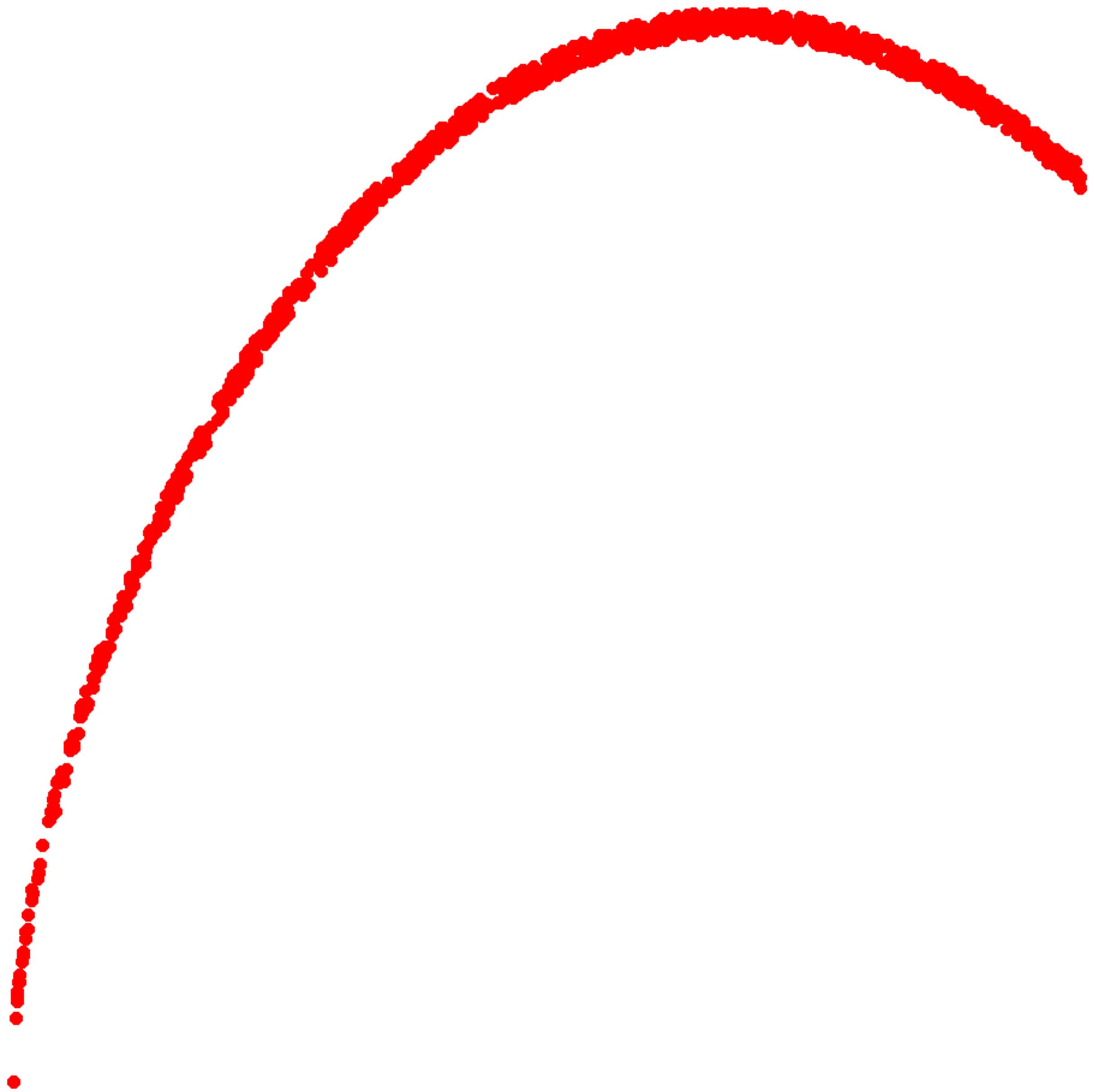
0.72687



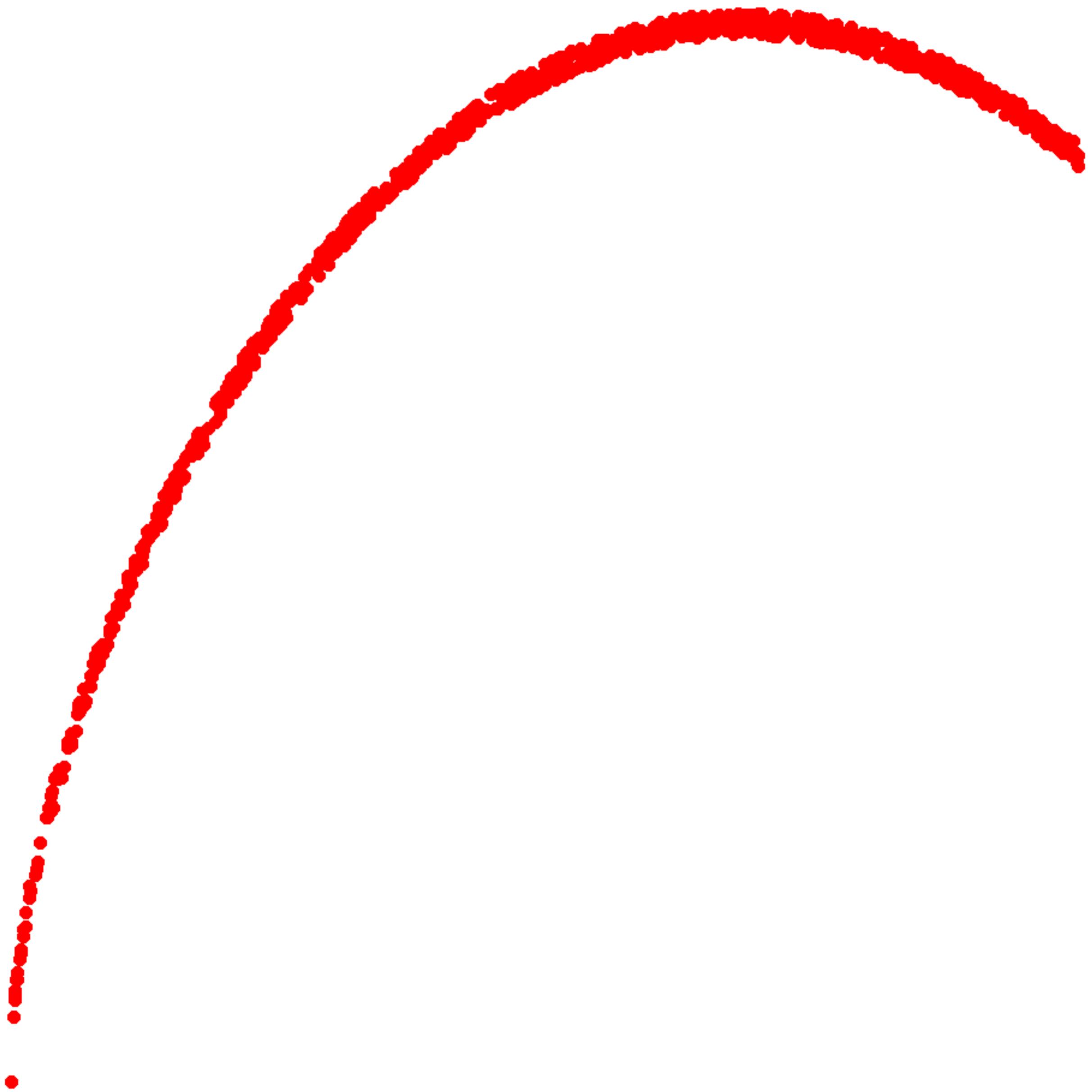
0.75579



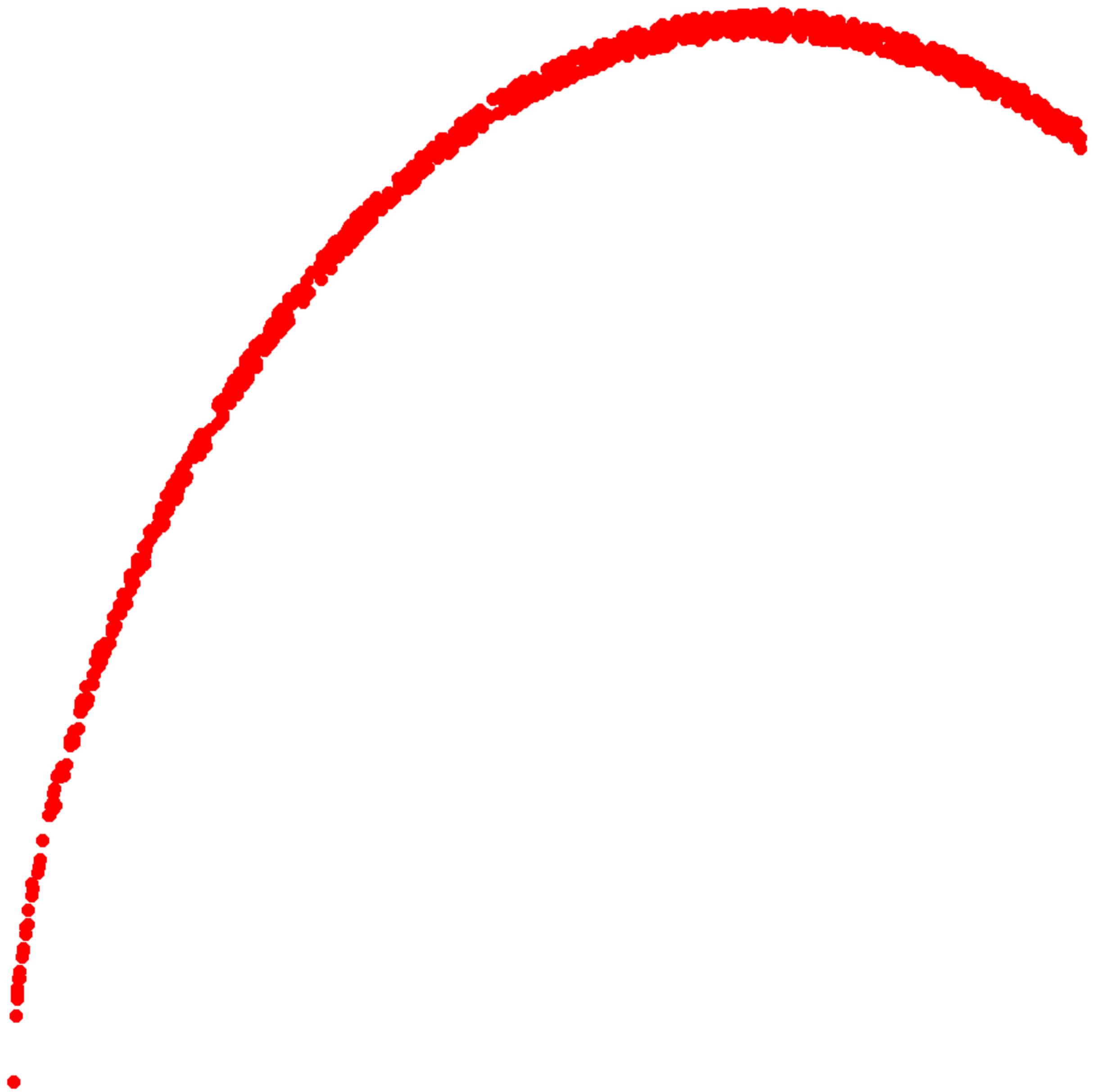
0.77898



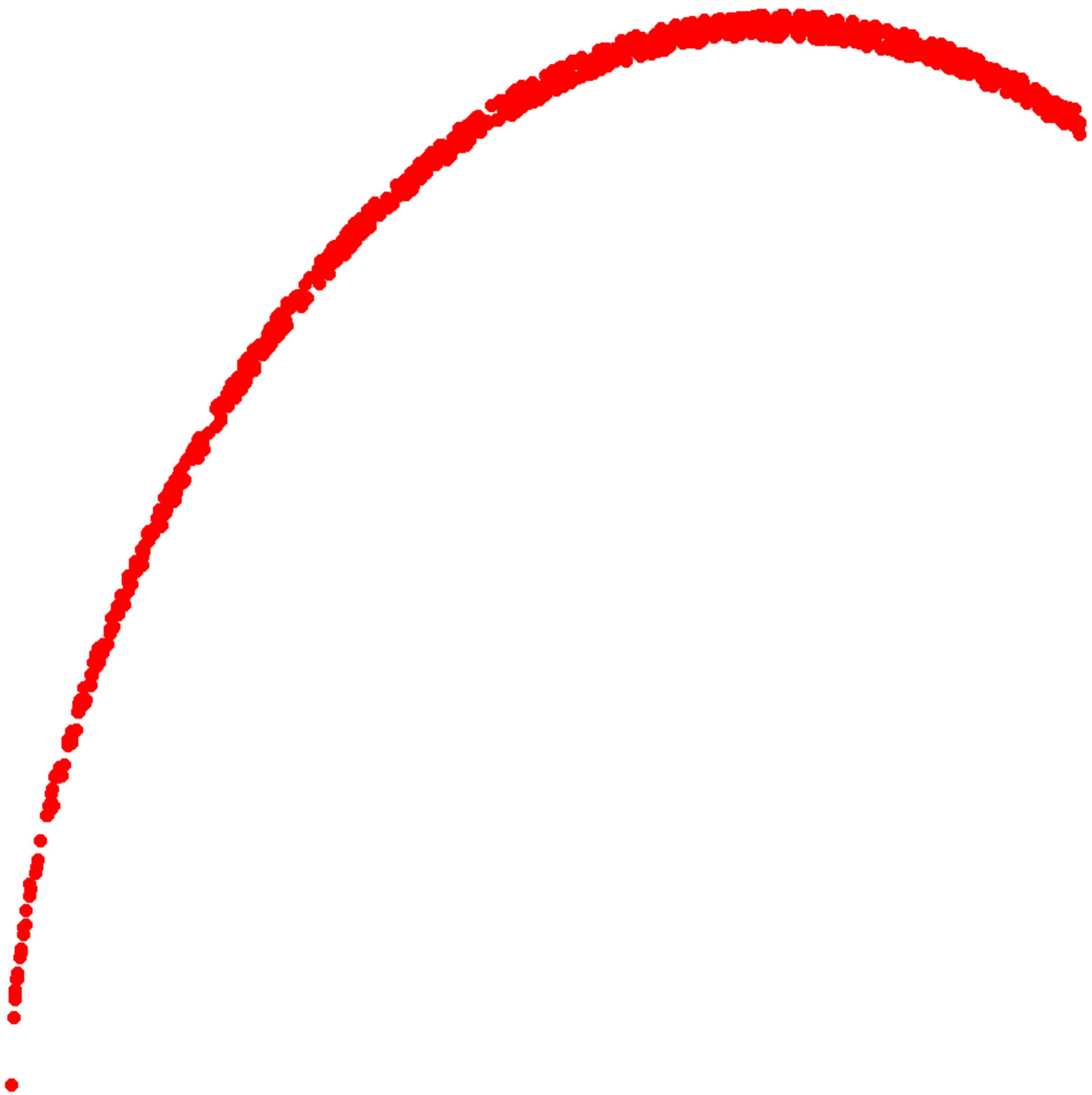
0.79785



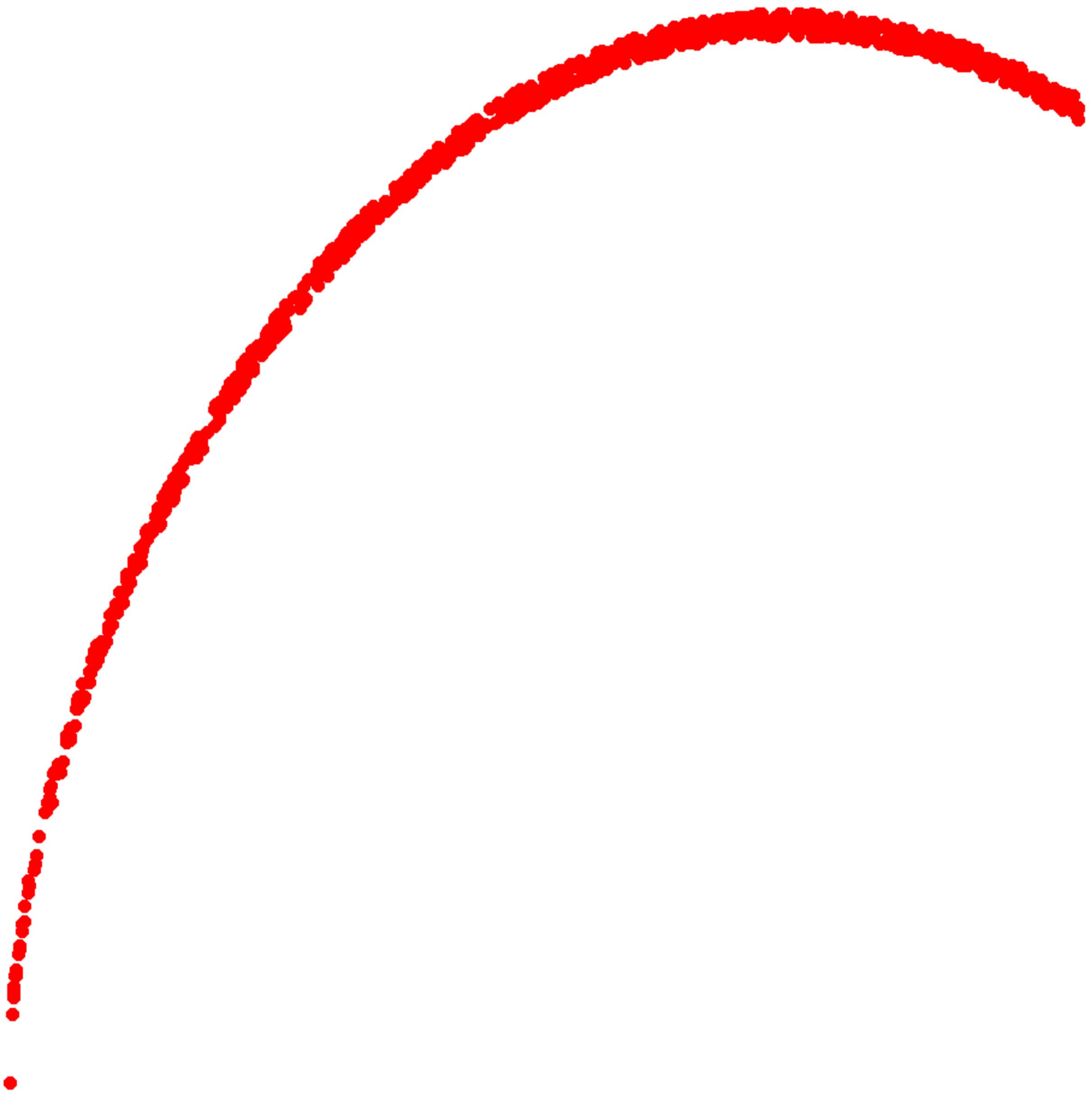
0.81344



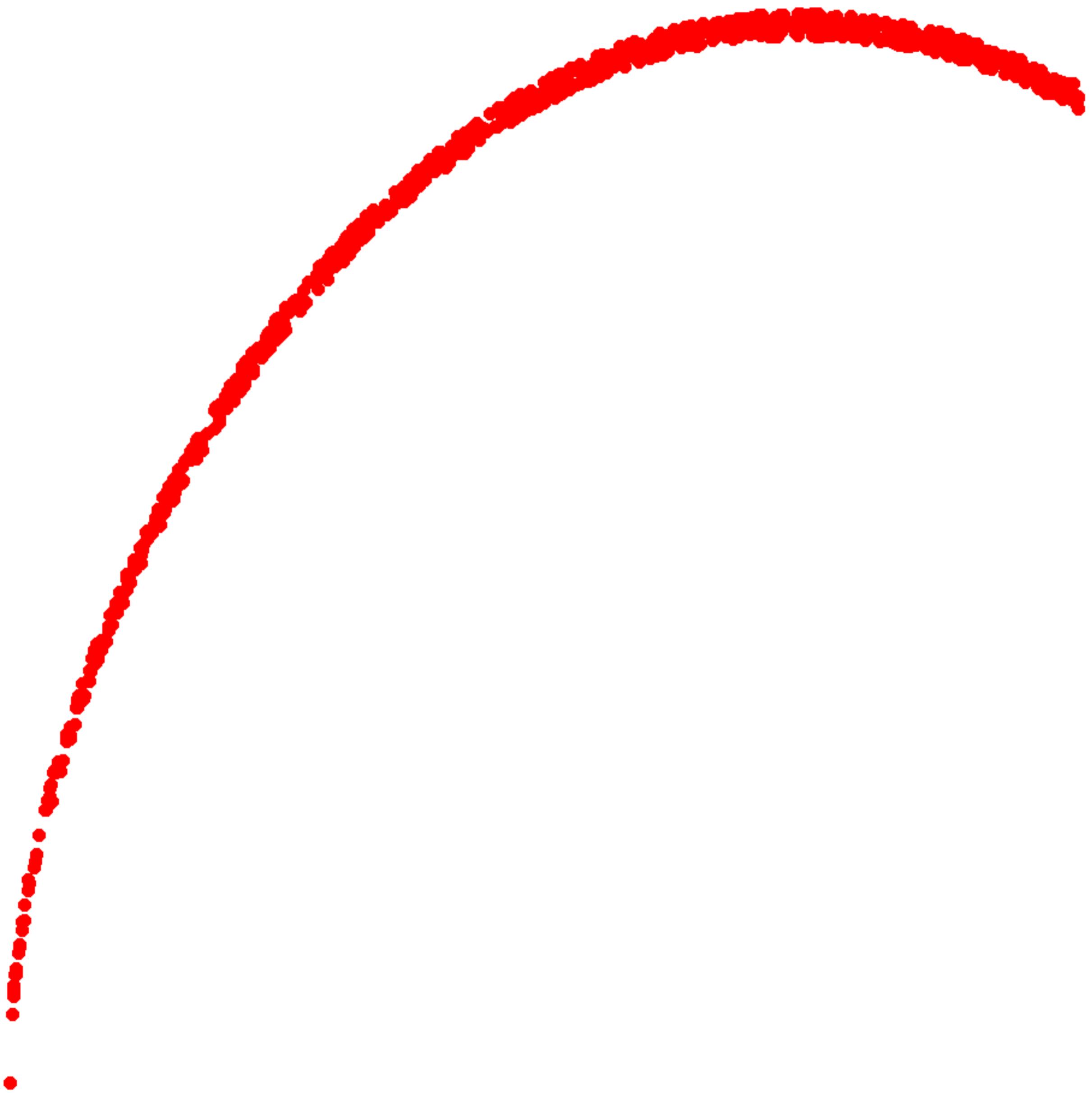
0.82647



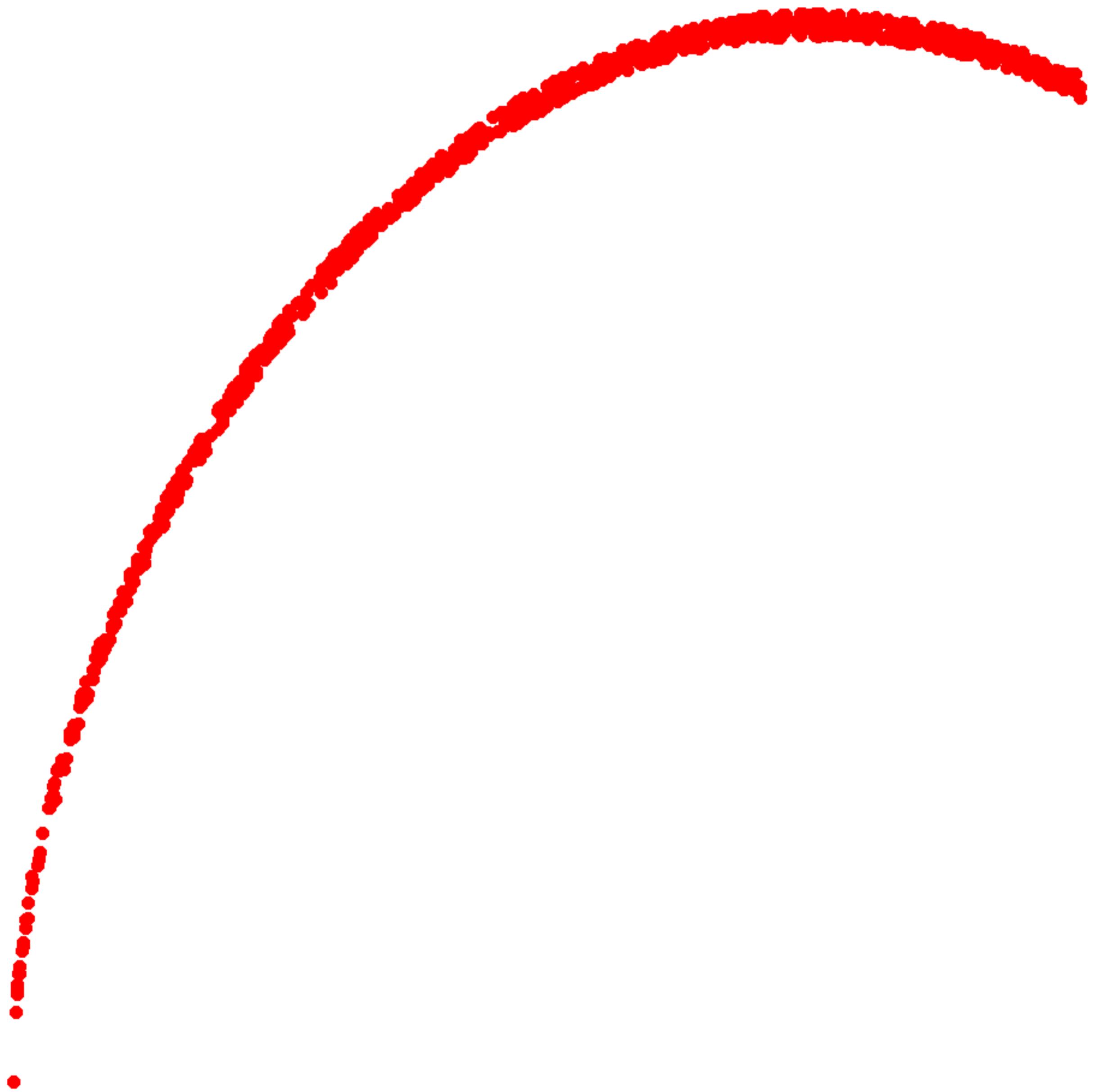
0.83749



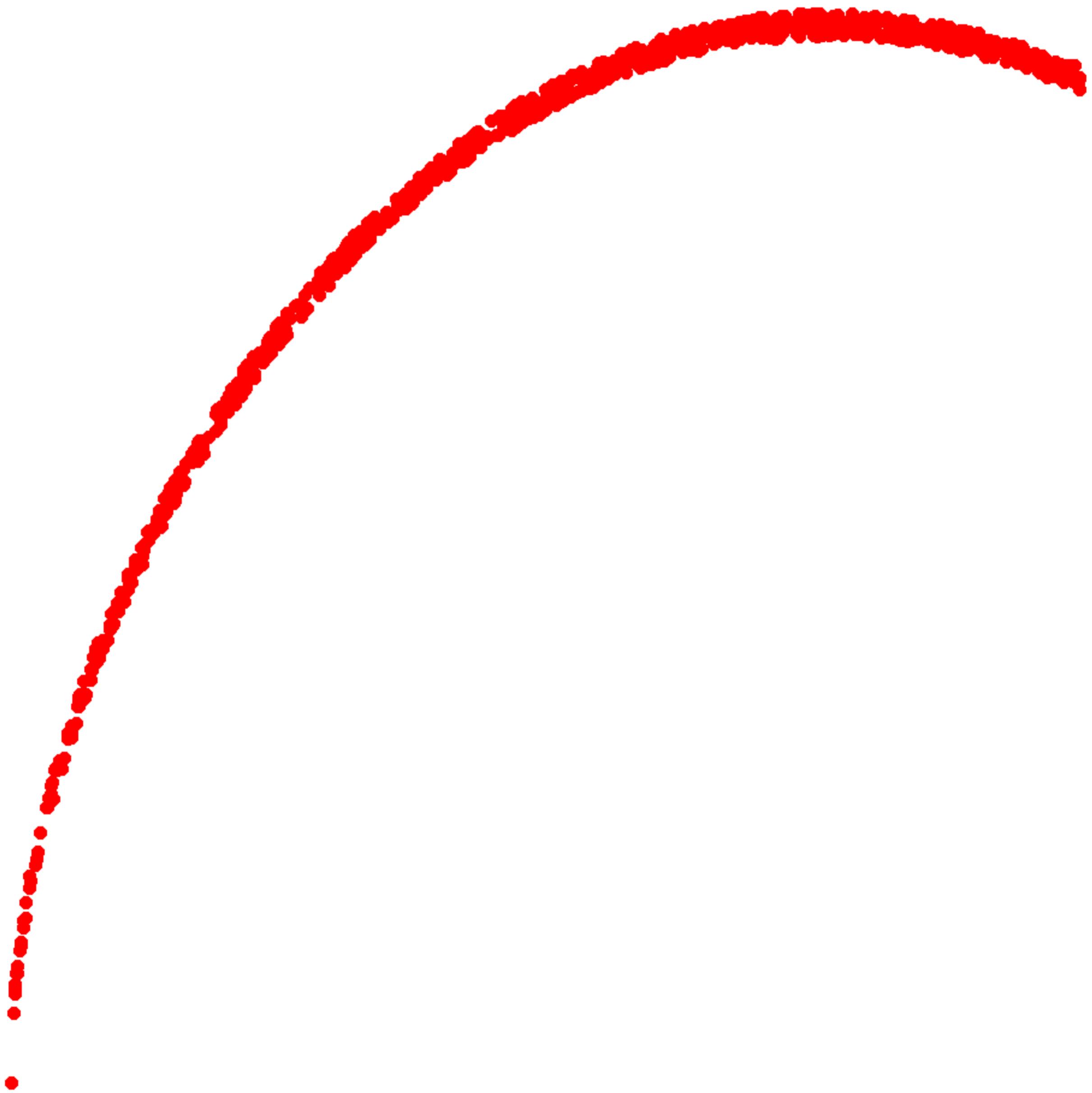
0.84691



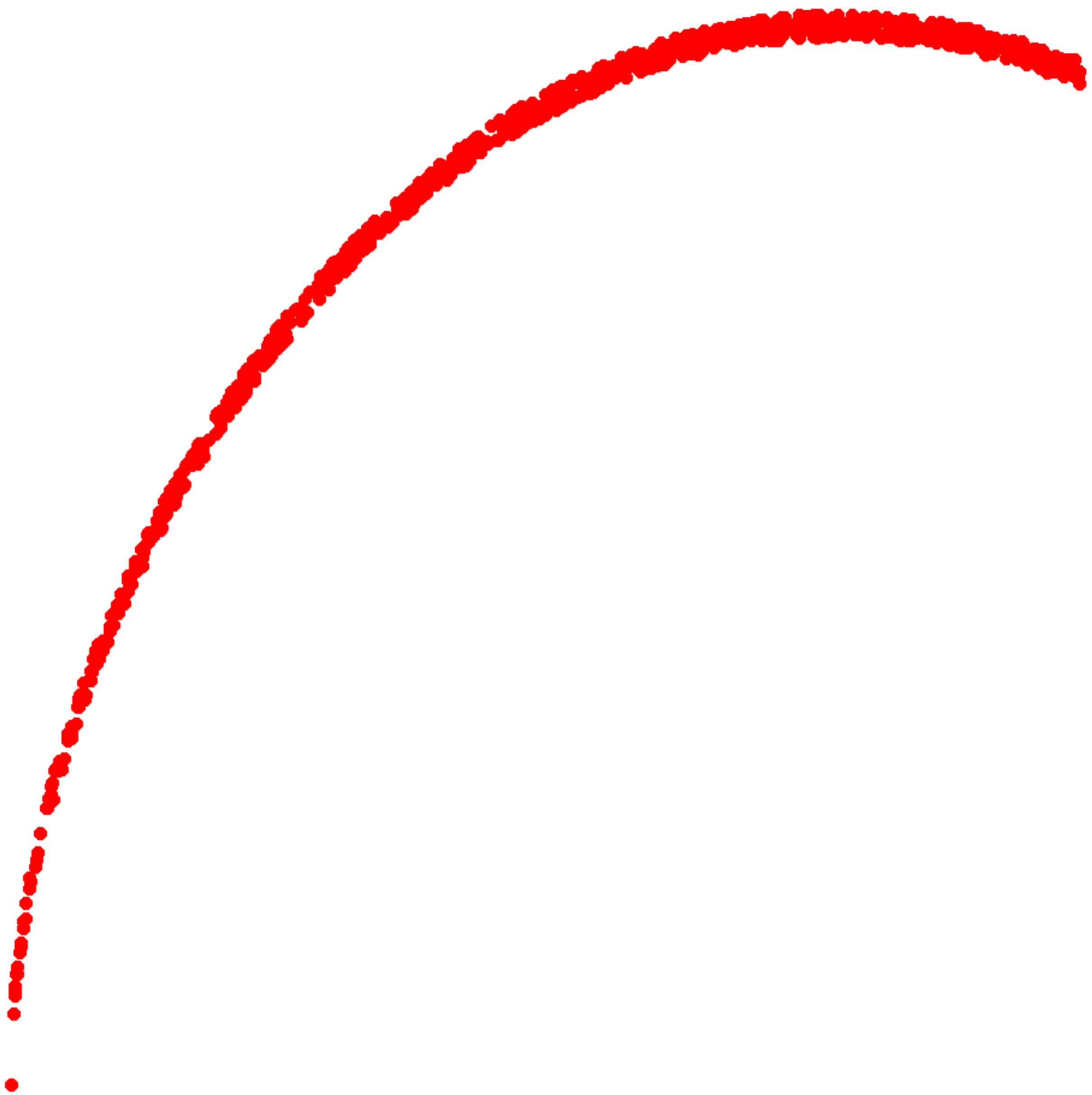
0.85504



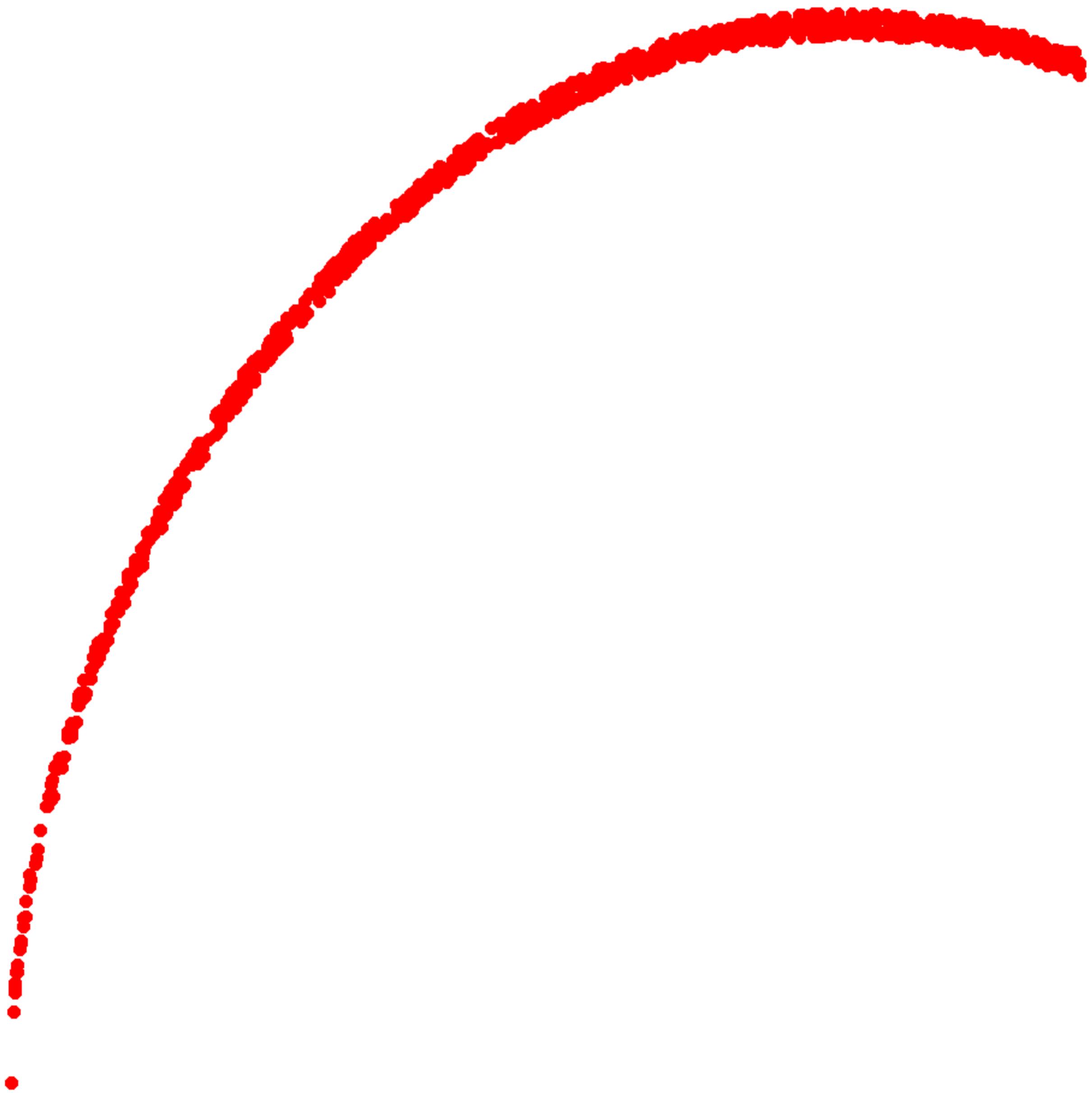
0.86211



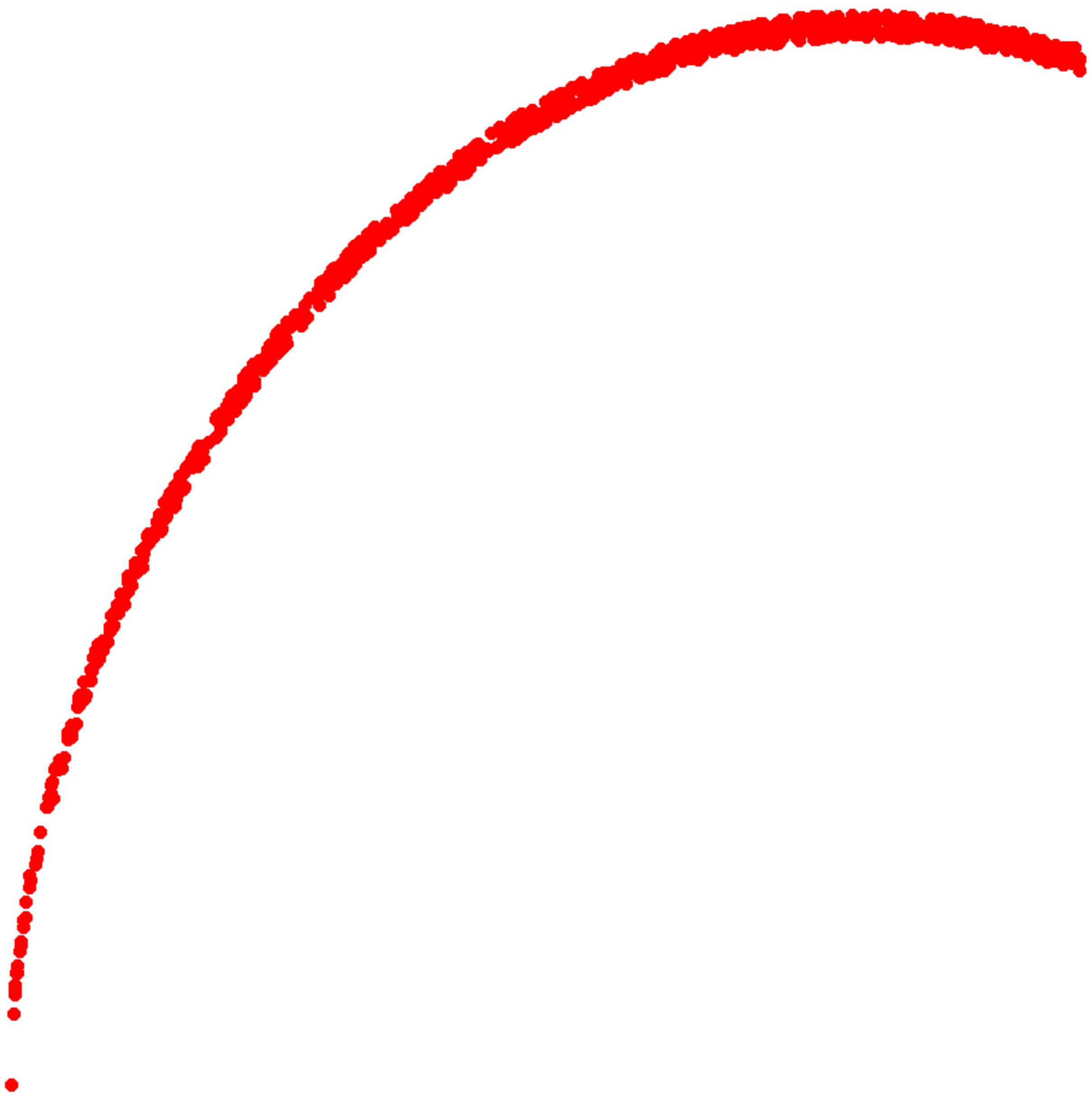
0.8683



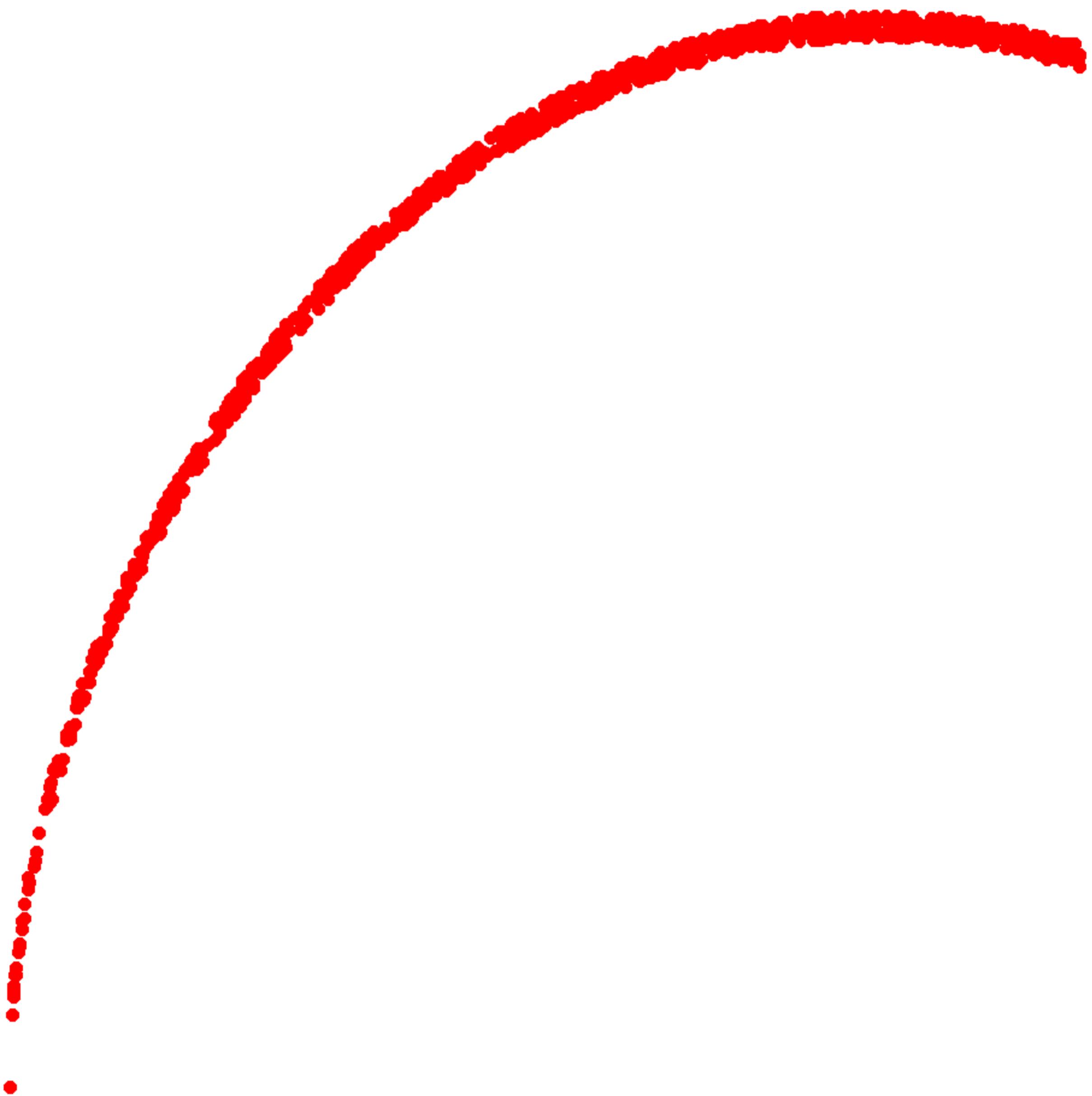
0.87377



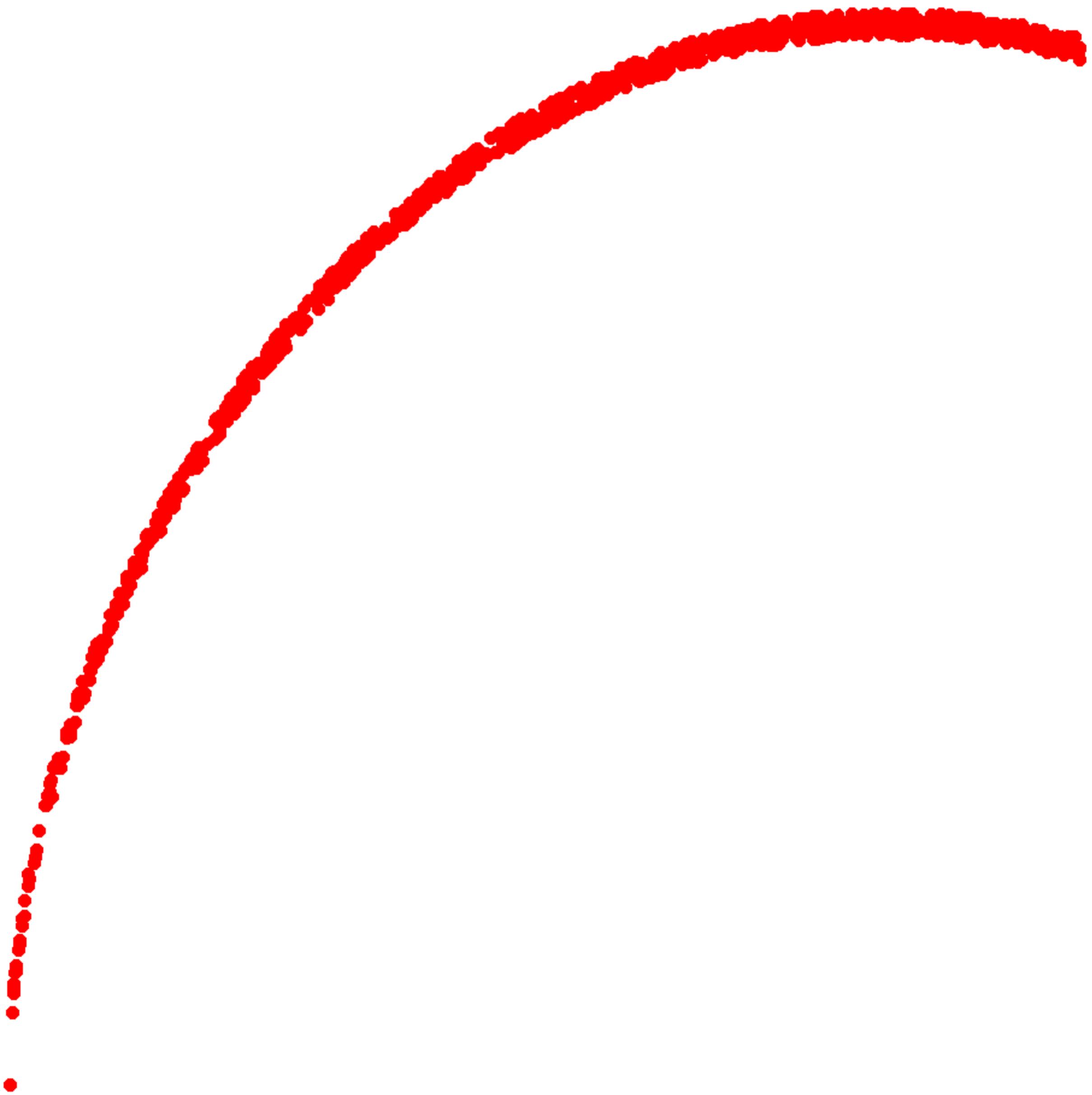
0.87862



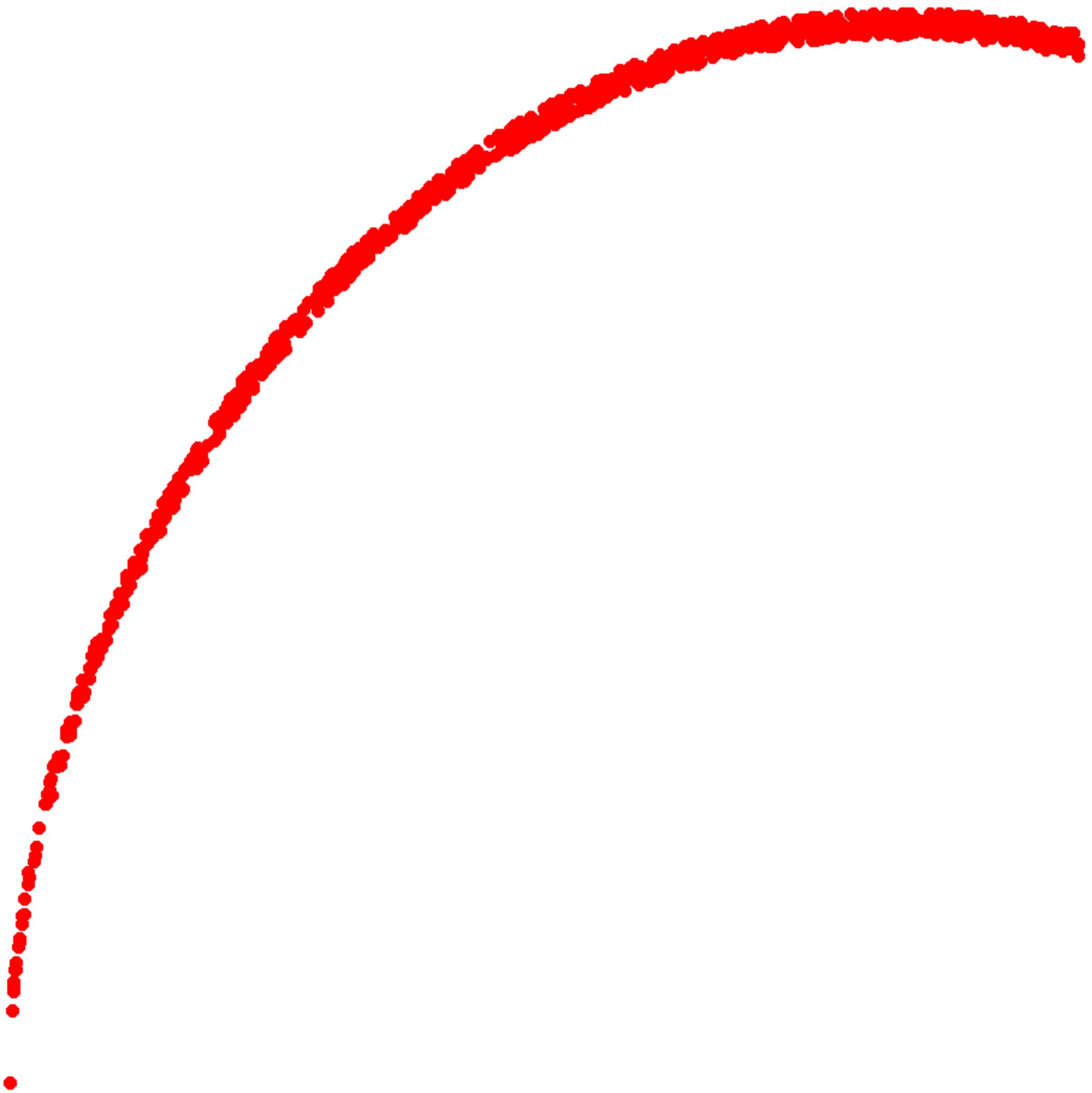
0.88296



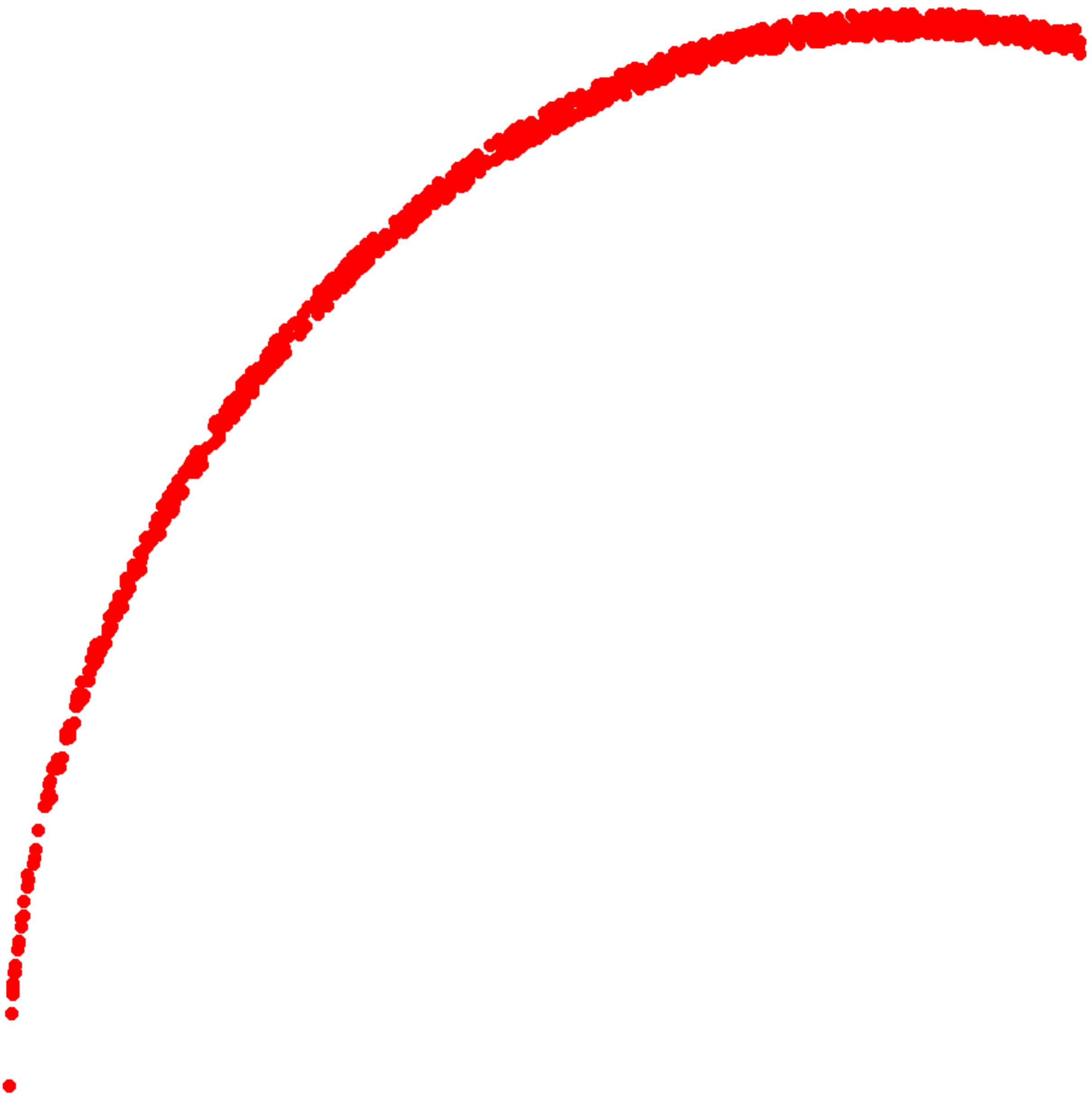
0.88686



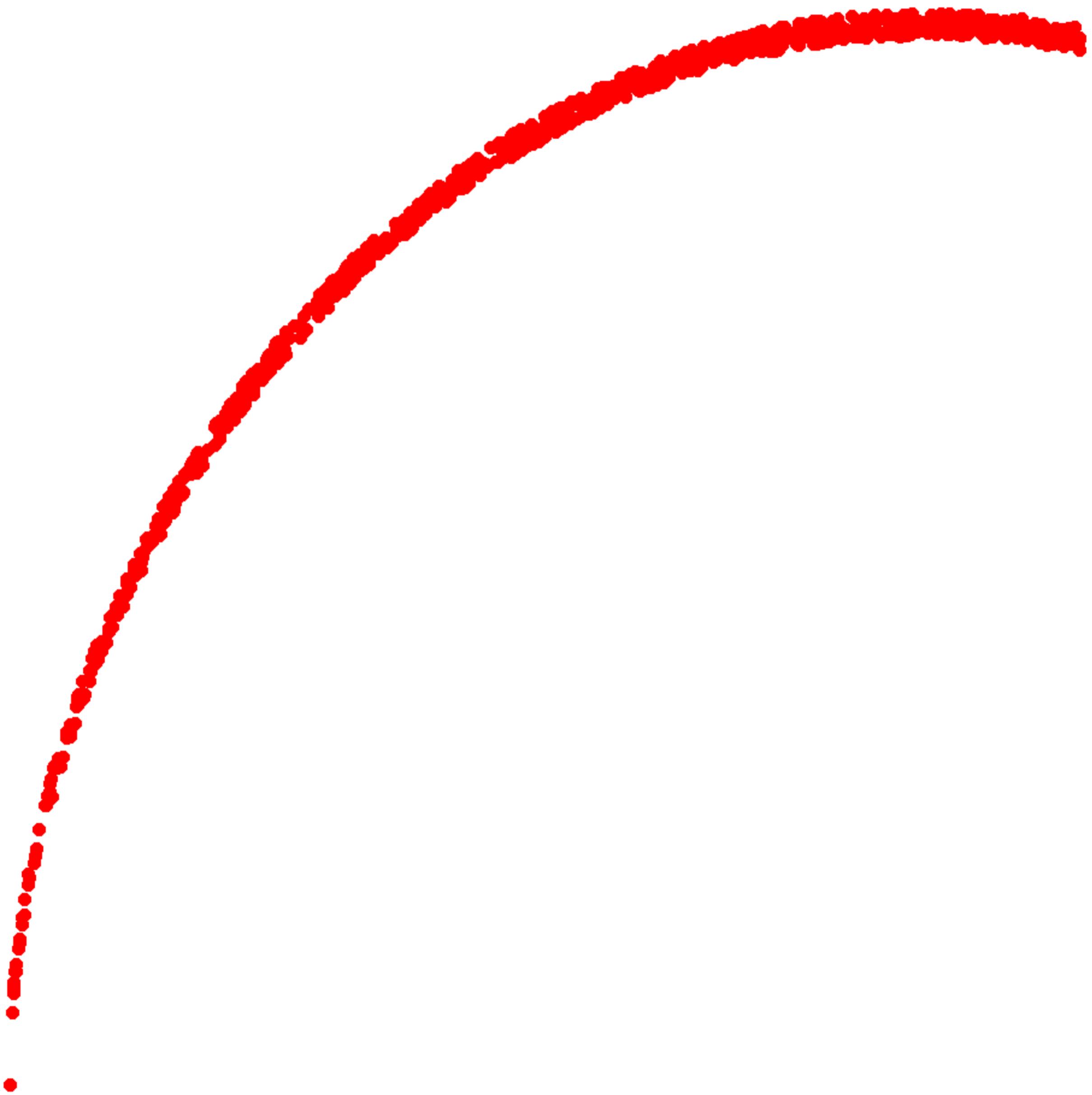
0.89037



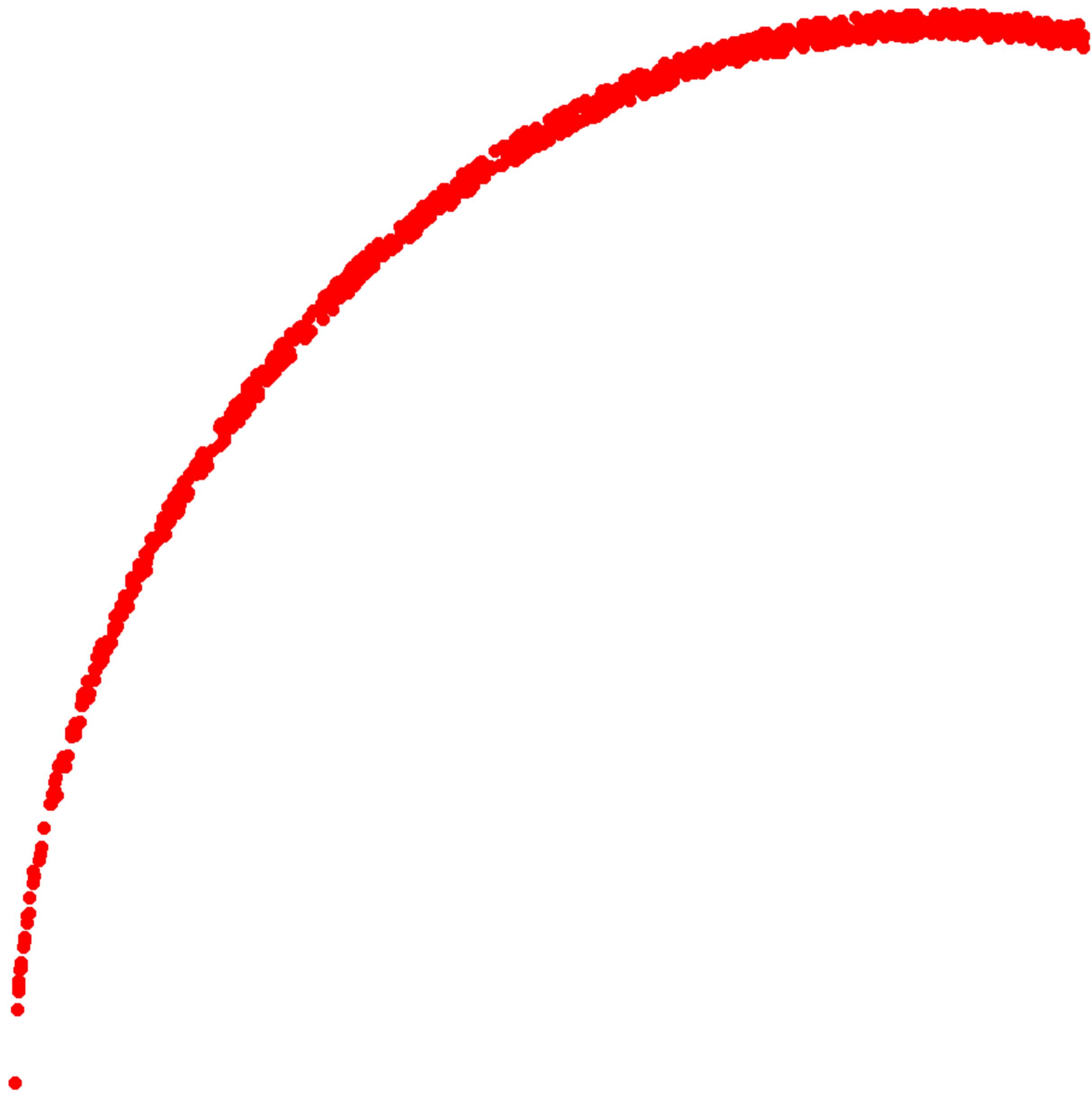
0.89356



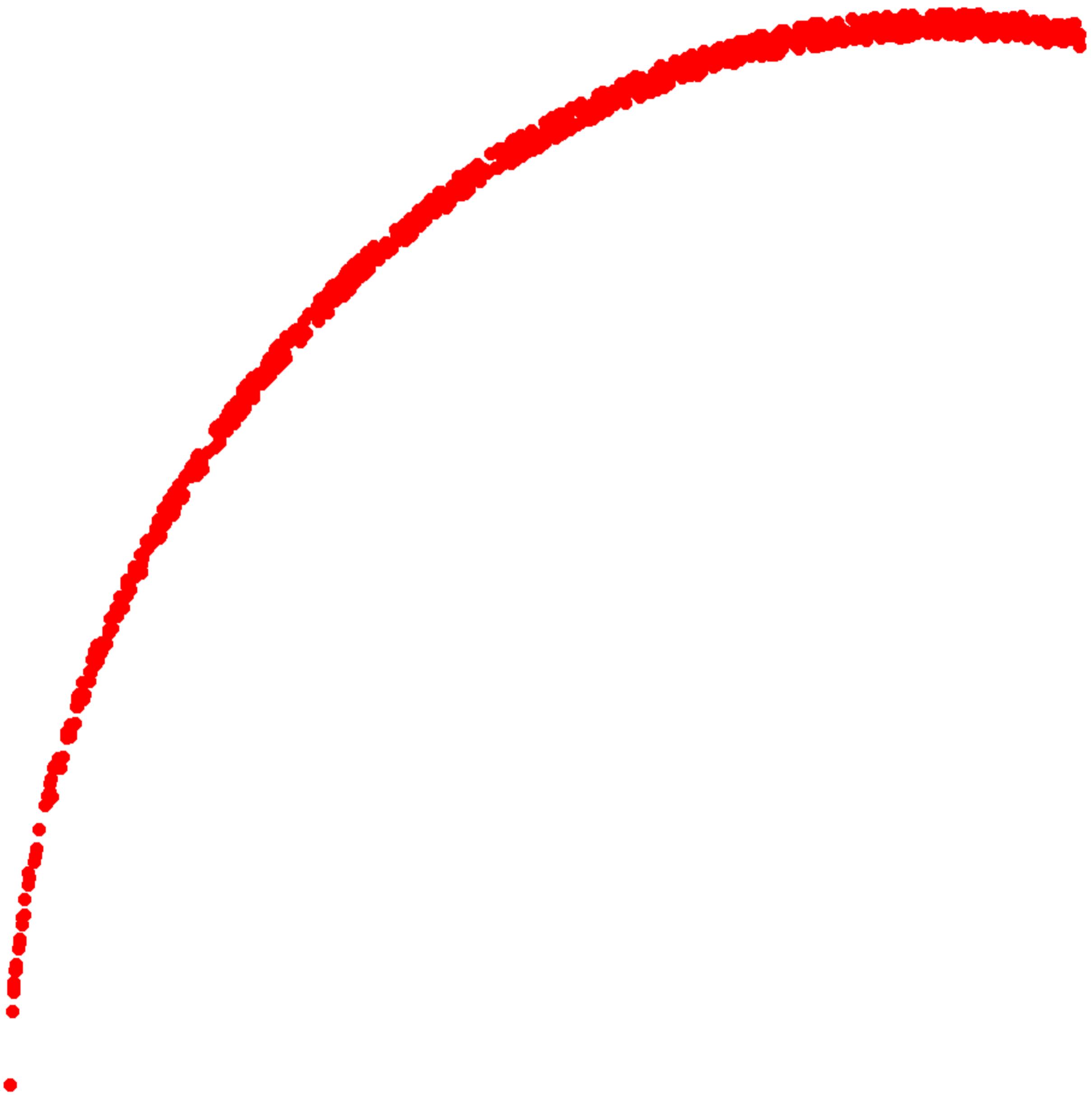
0.89646



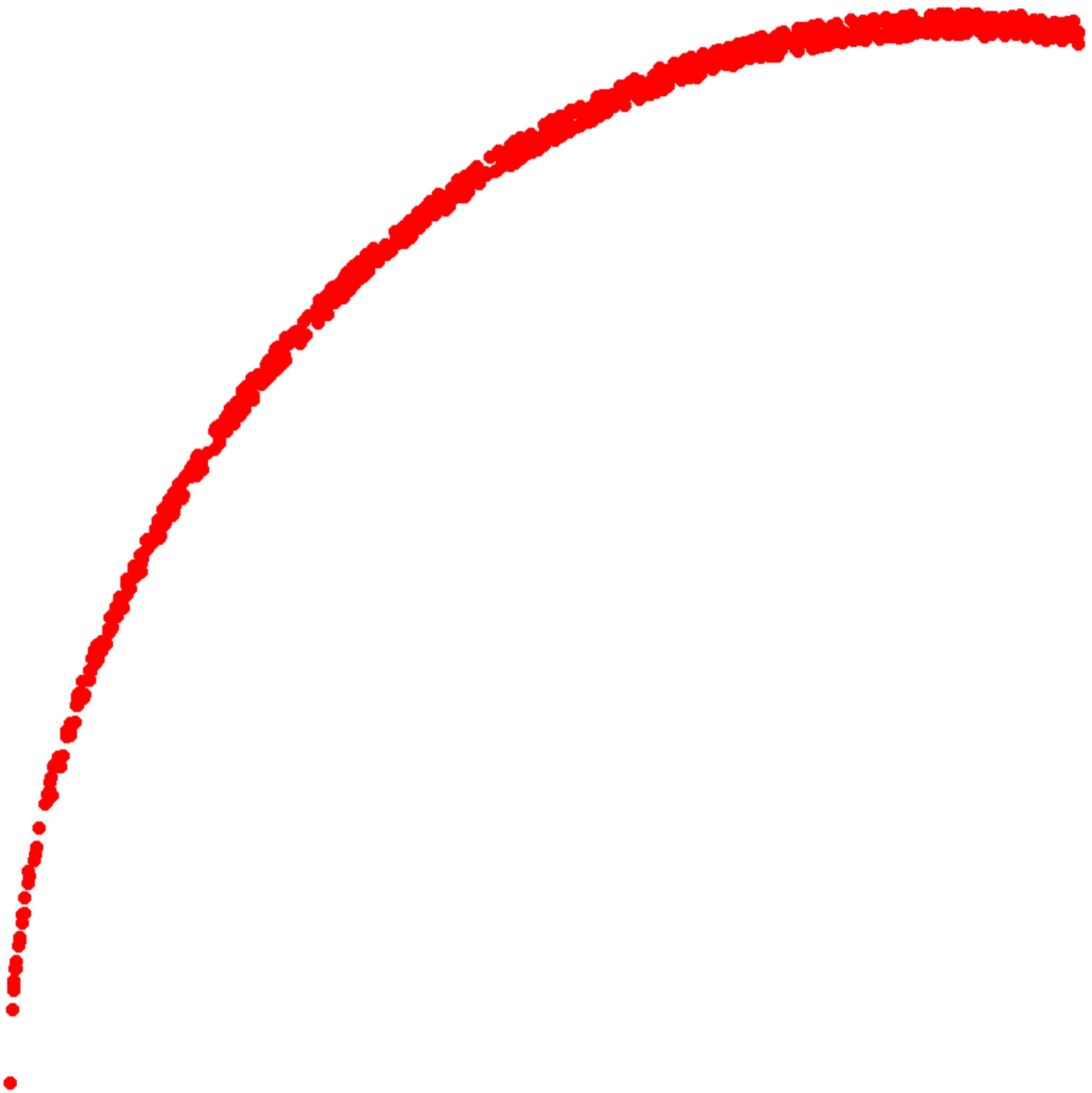
0.8991



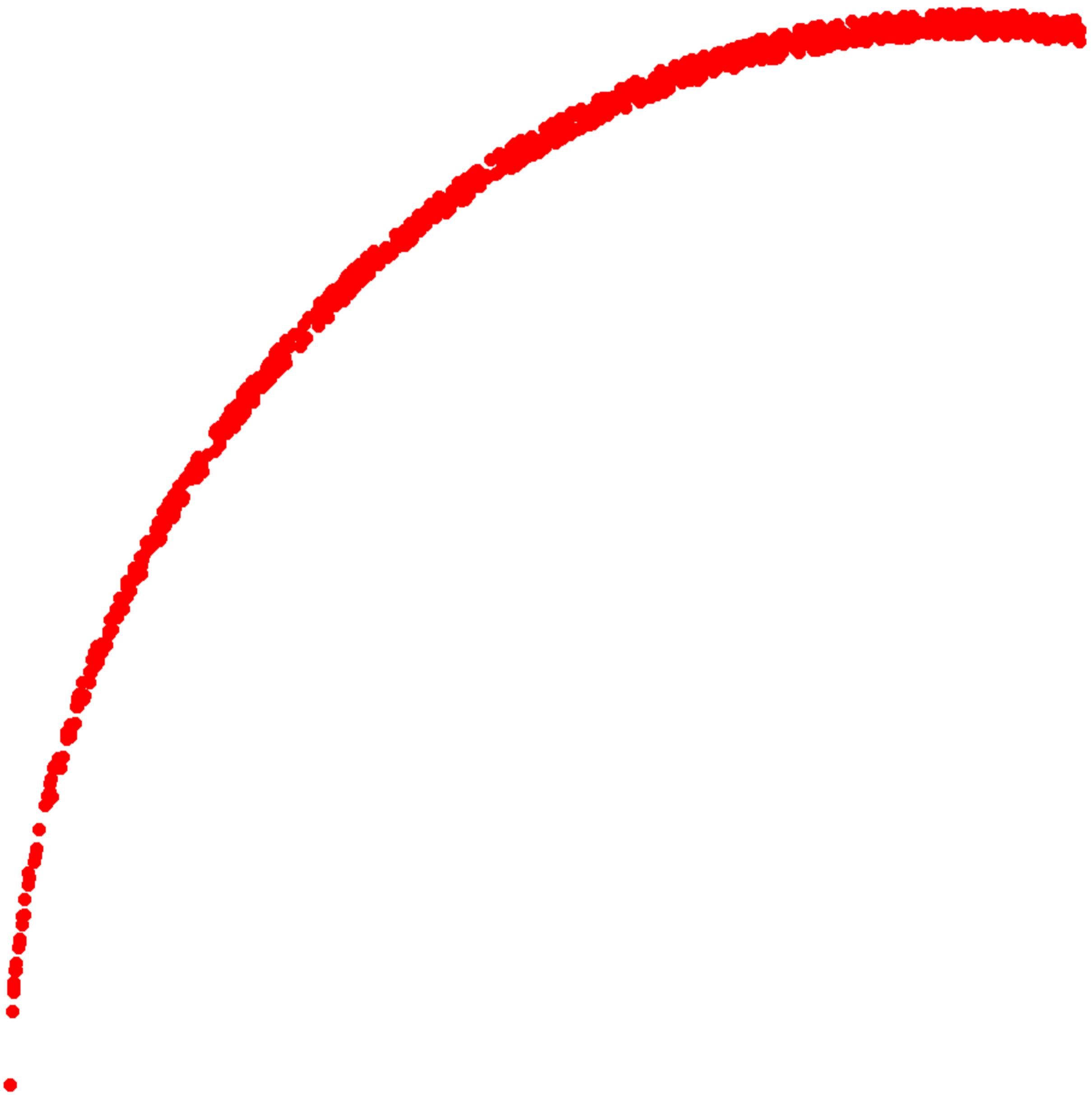
0.90153



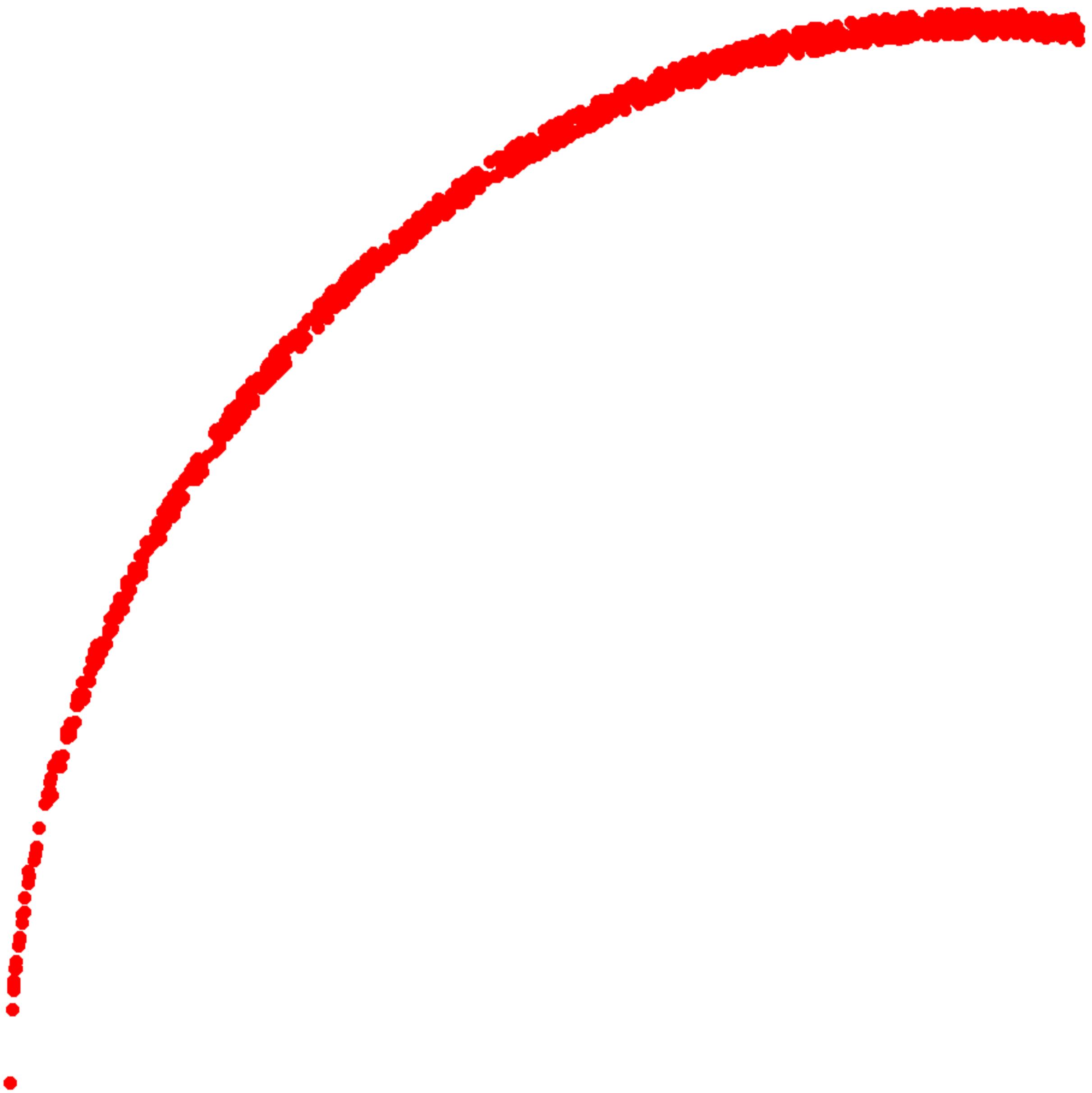
0.90376



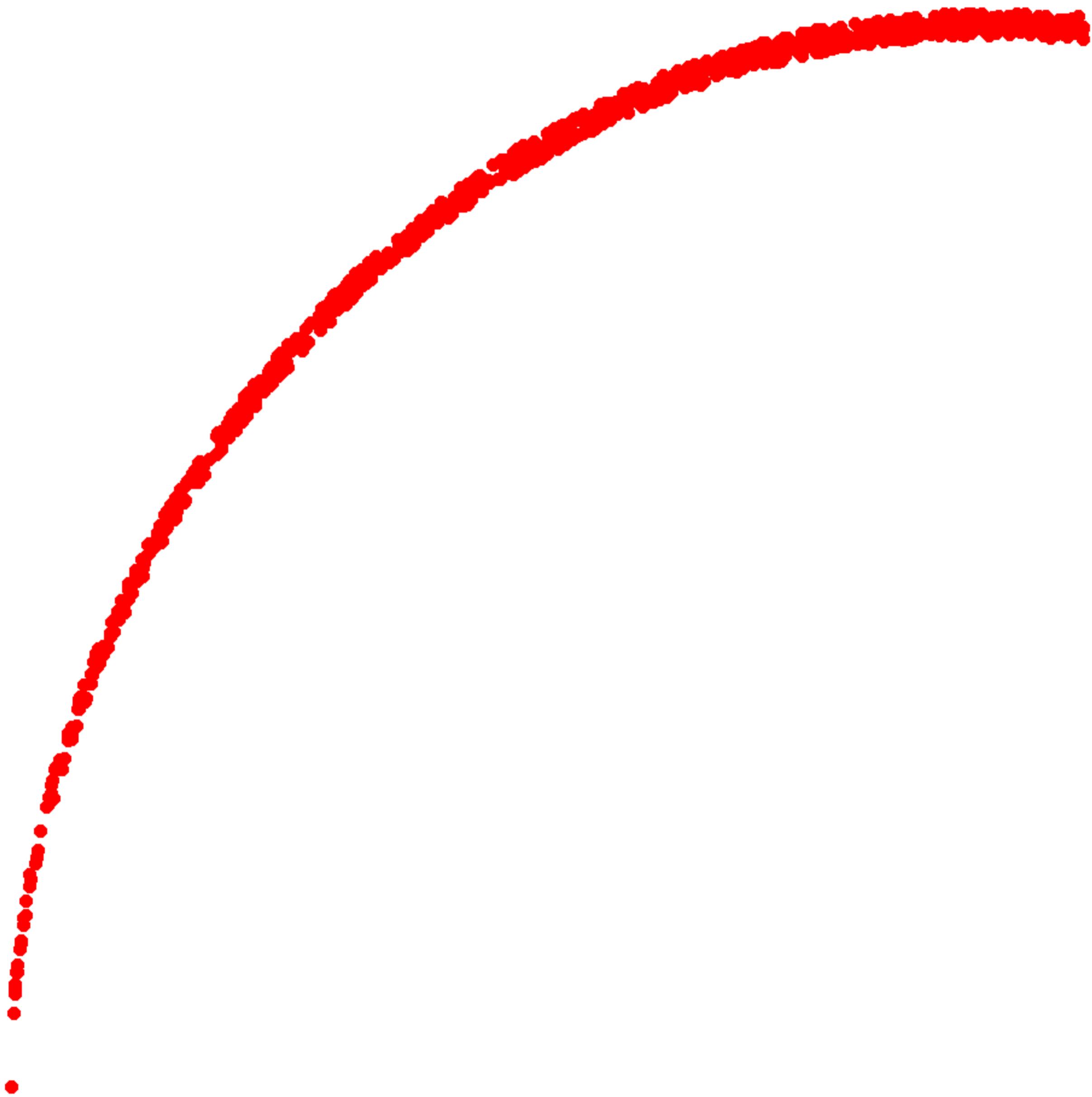
0.90582



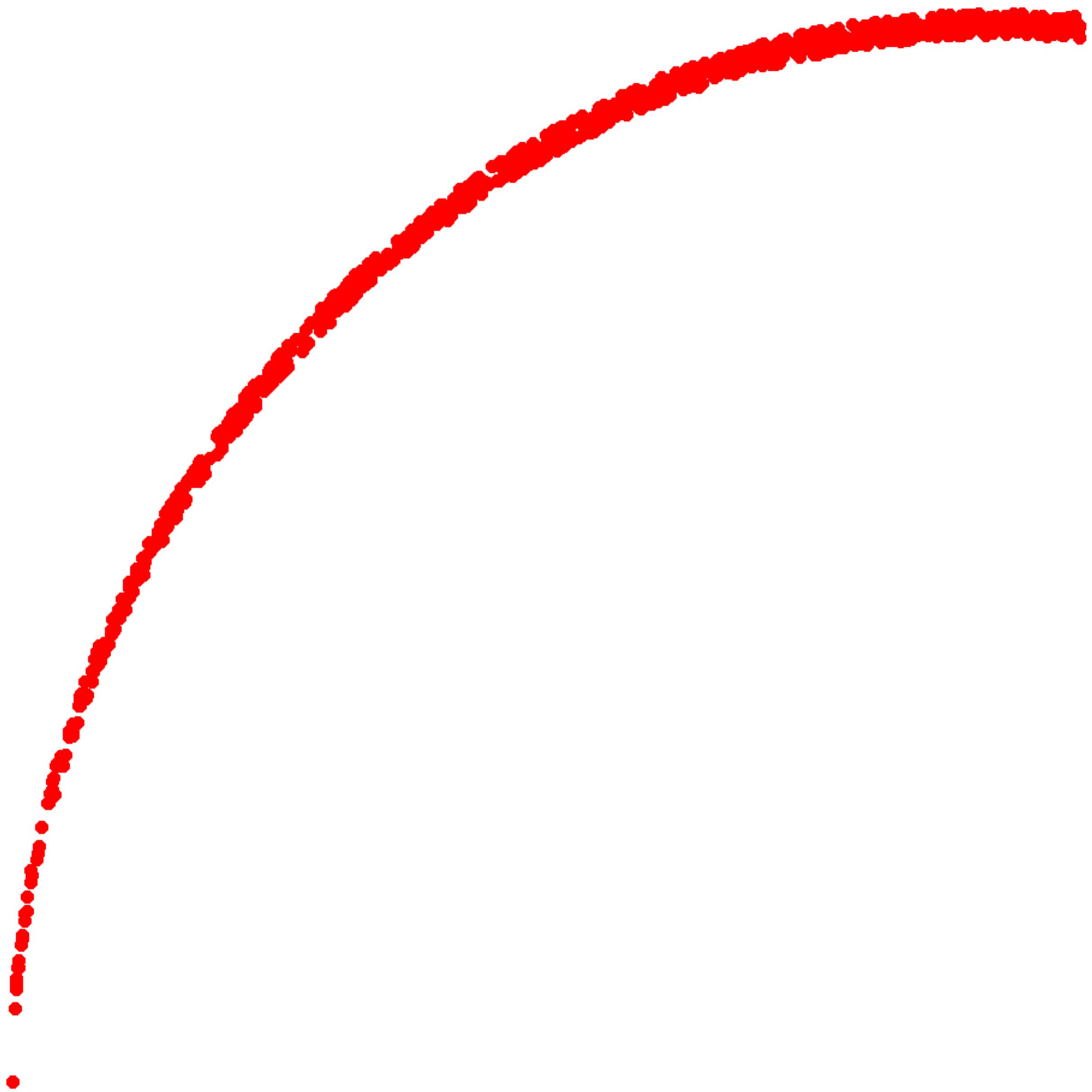
0.90773



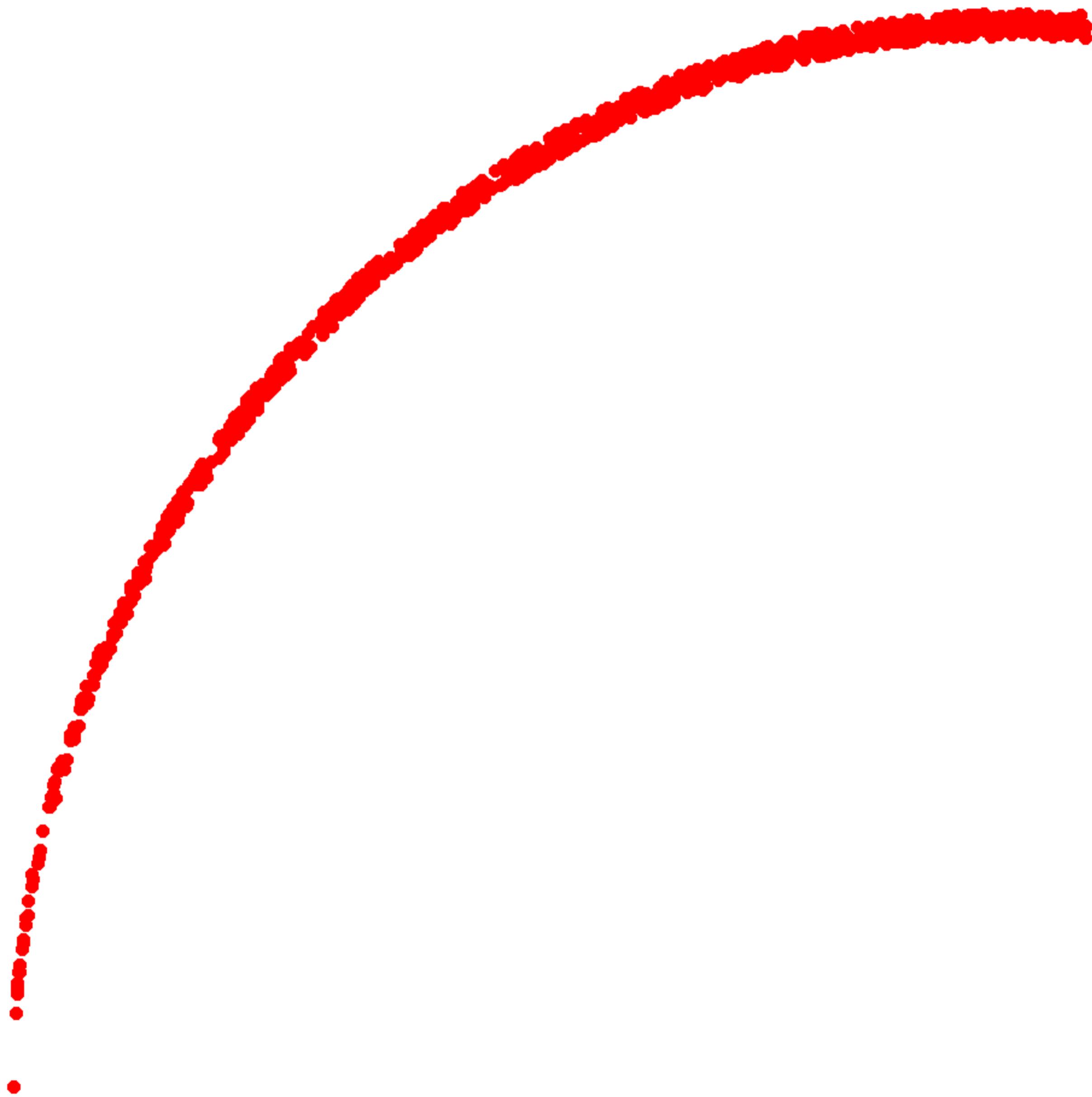
0.9095



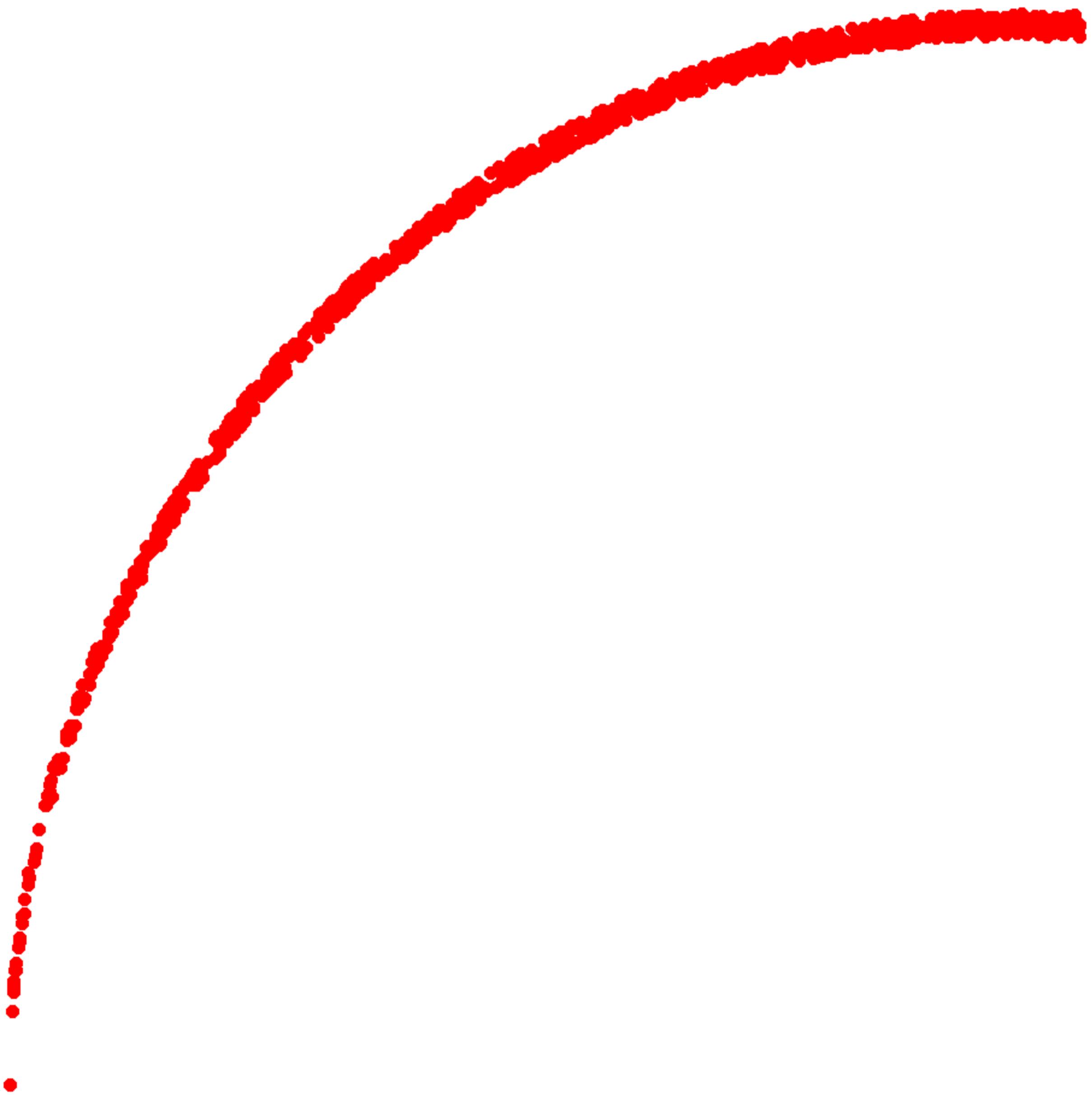
0.91114



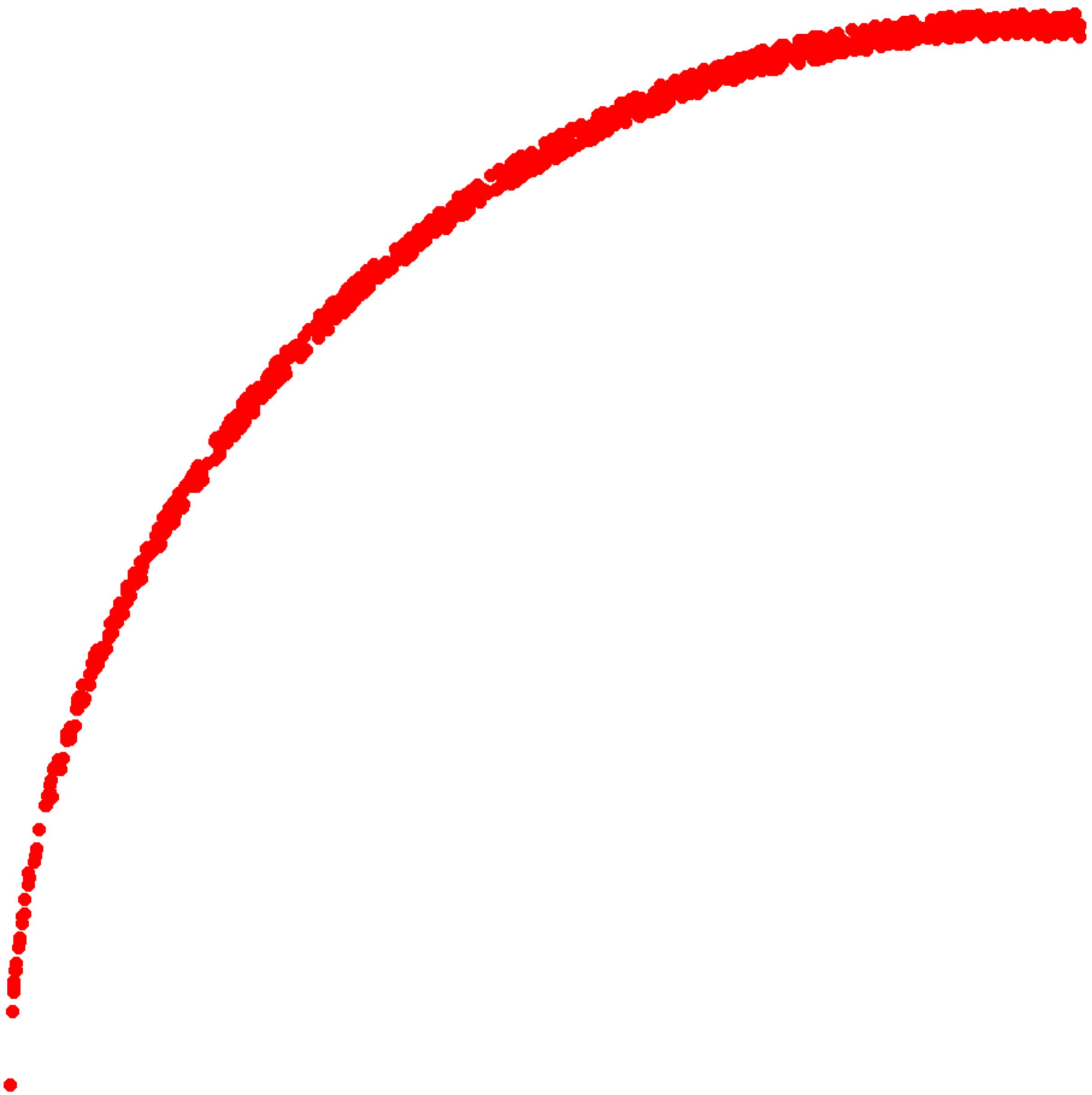
0.91268



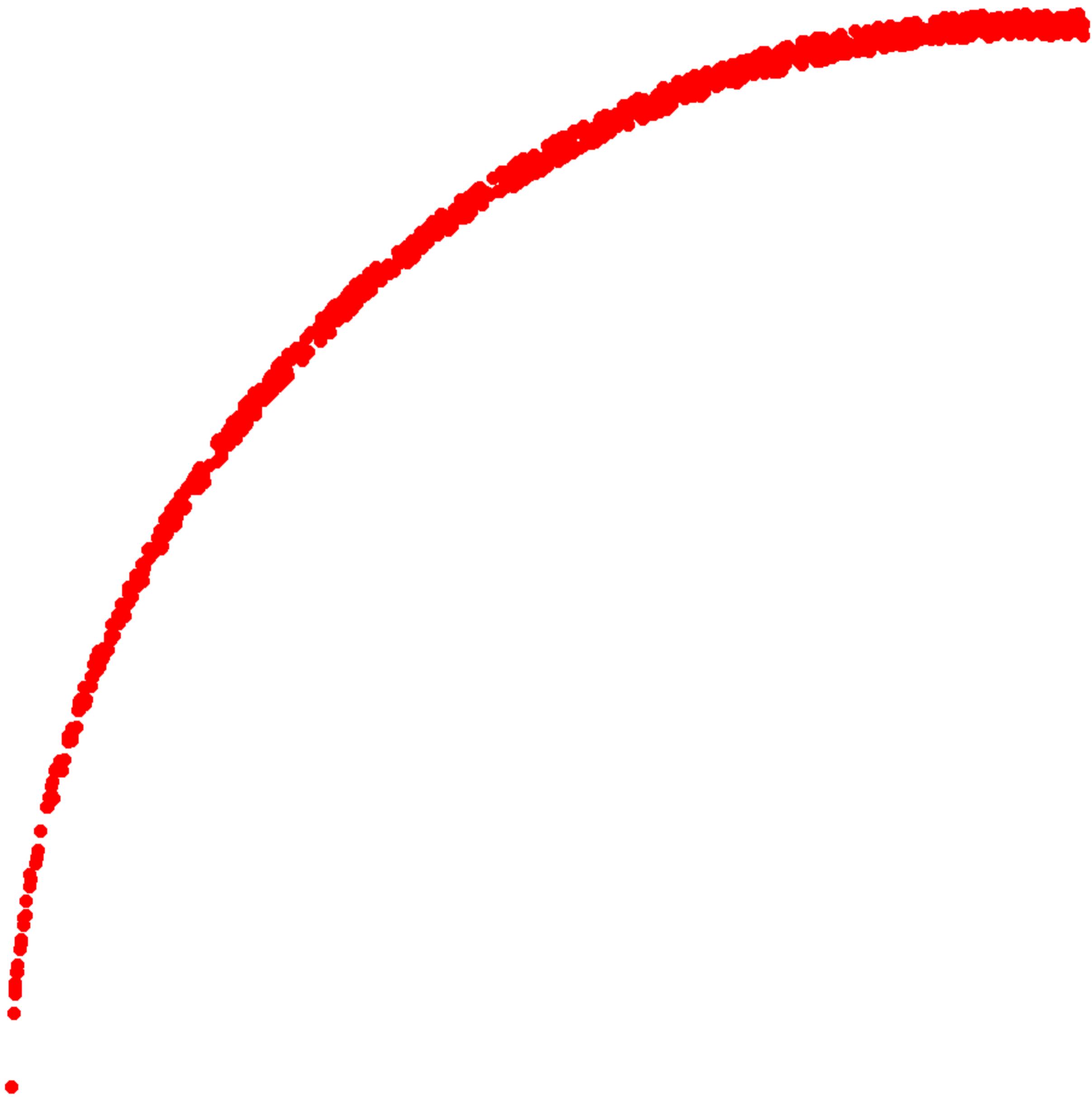
0.91411



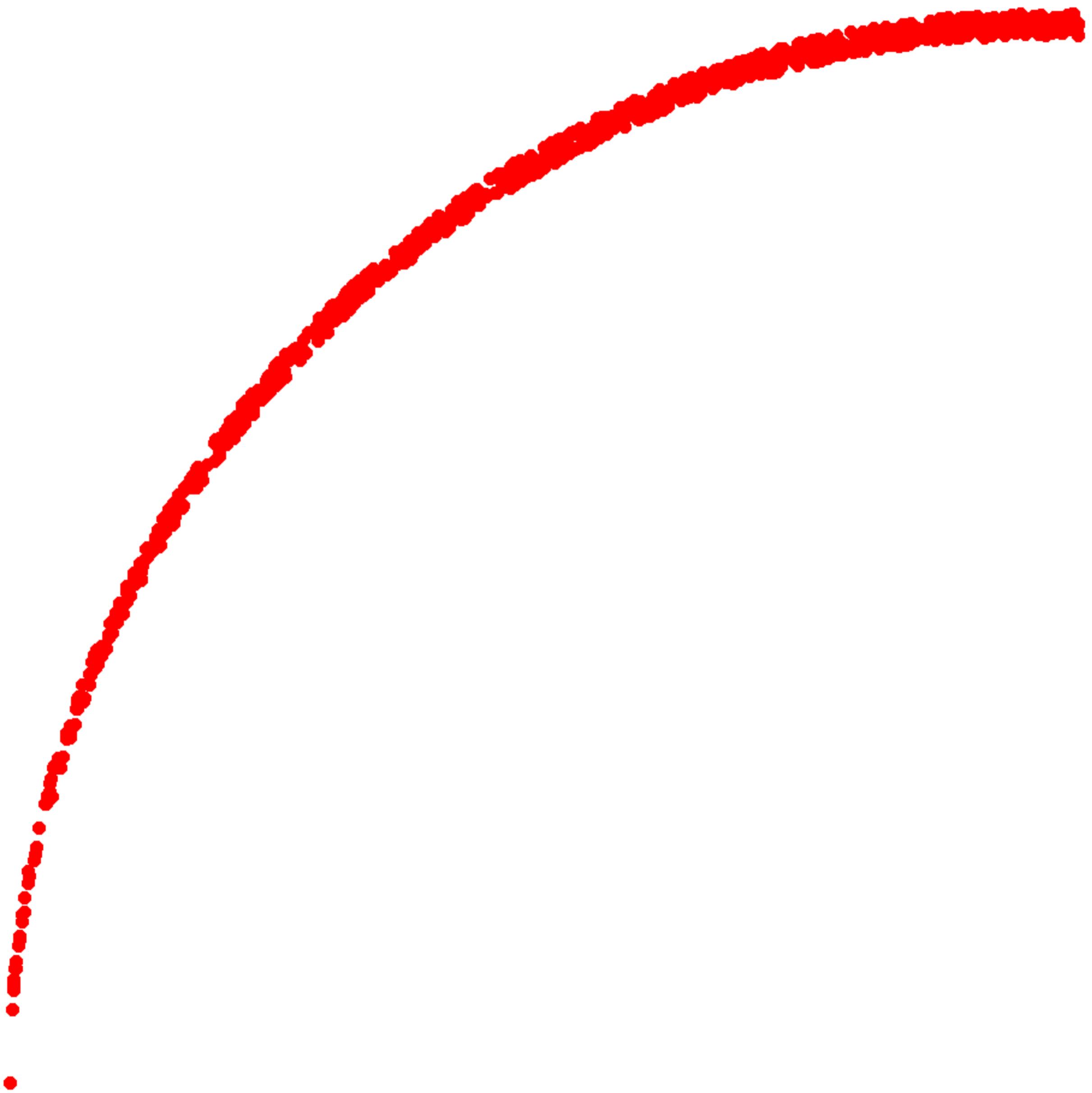
0.91545



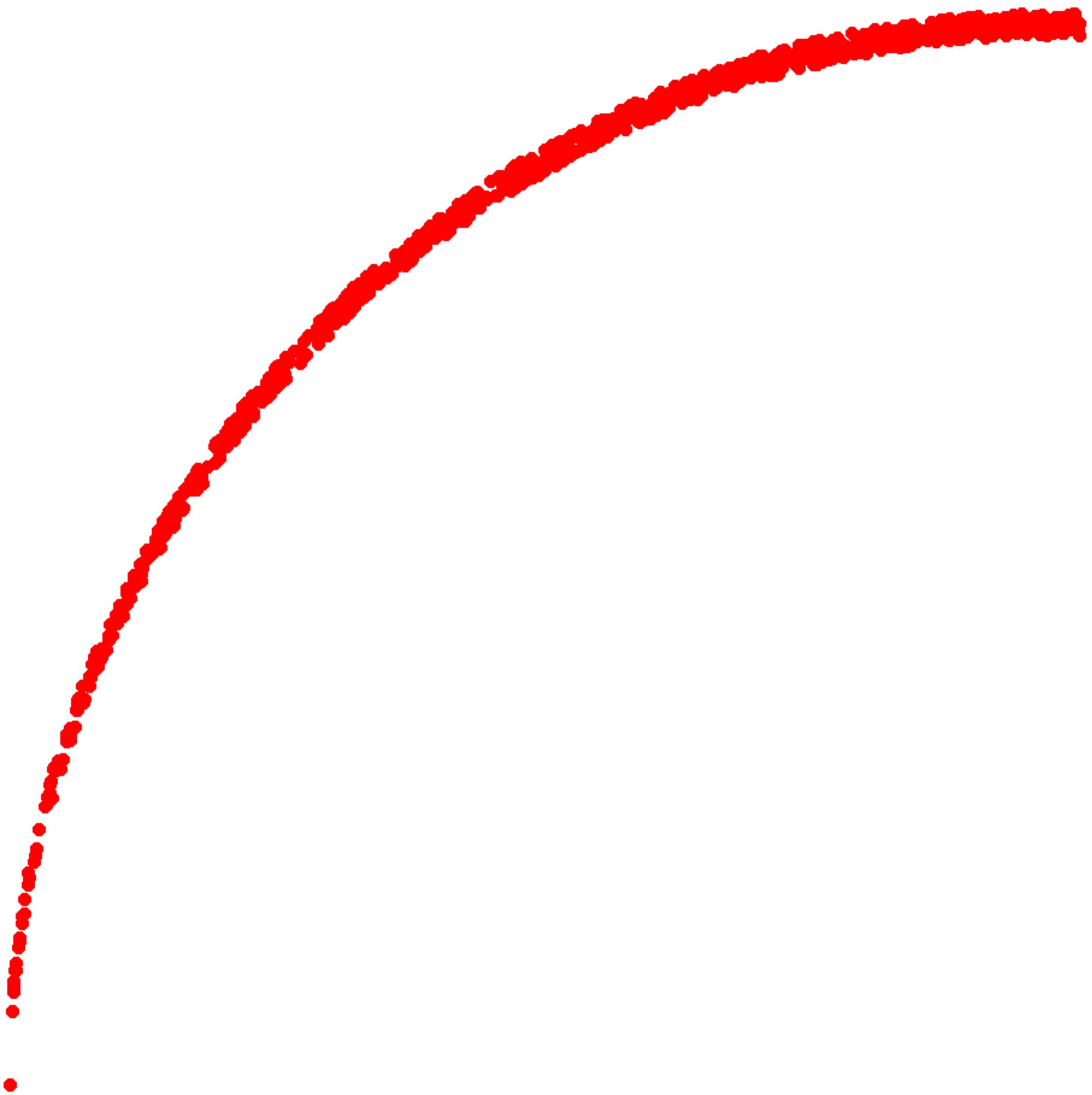
0.9167



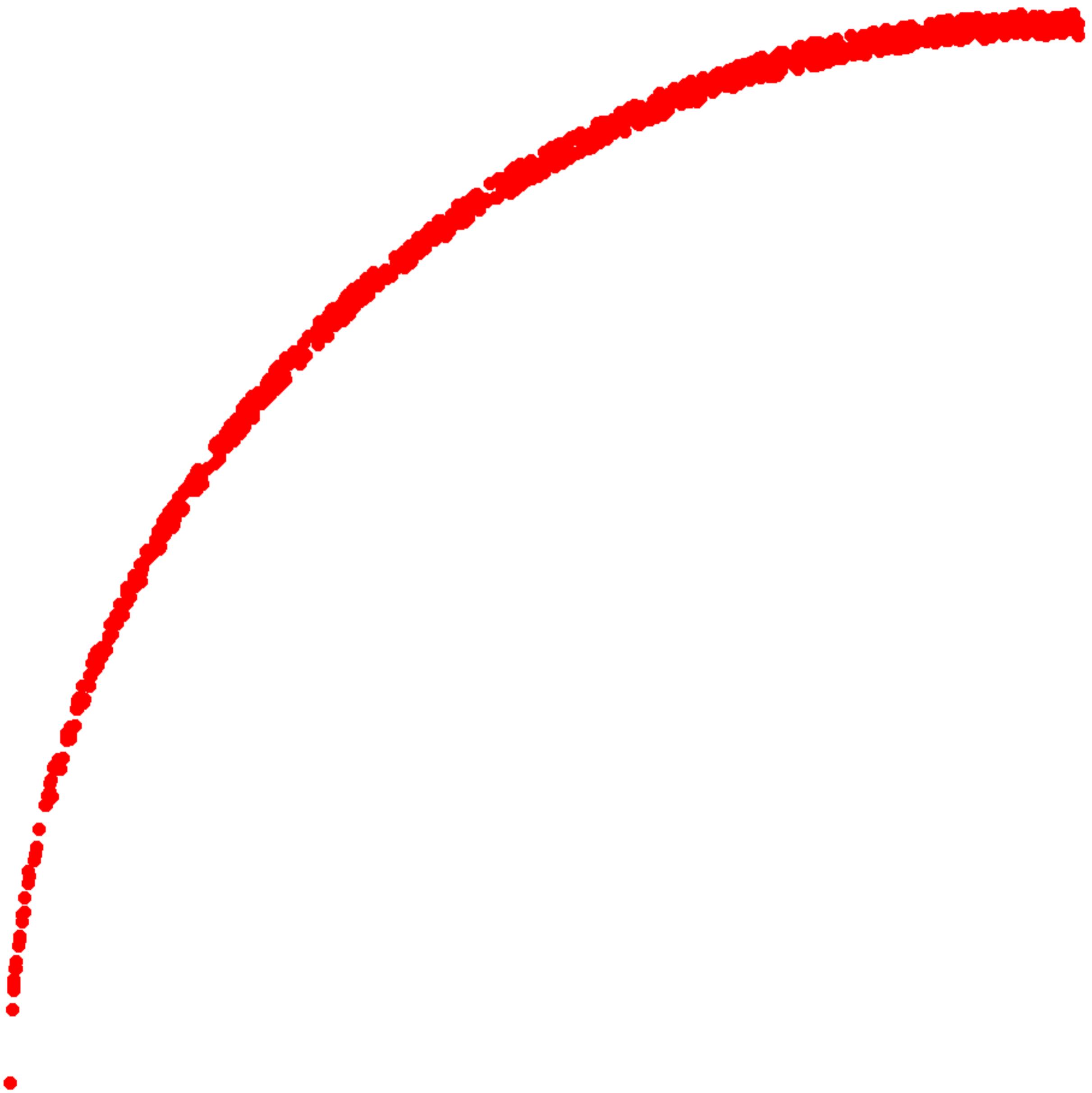
0.91789



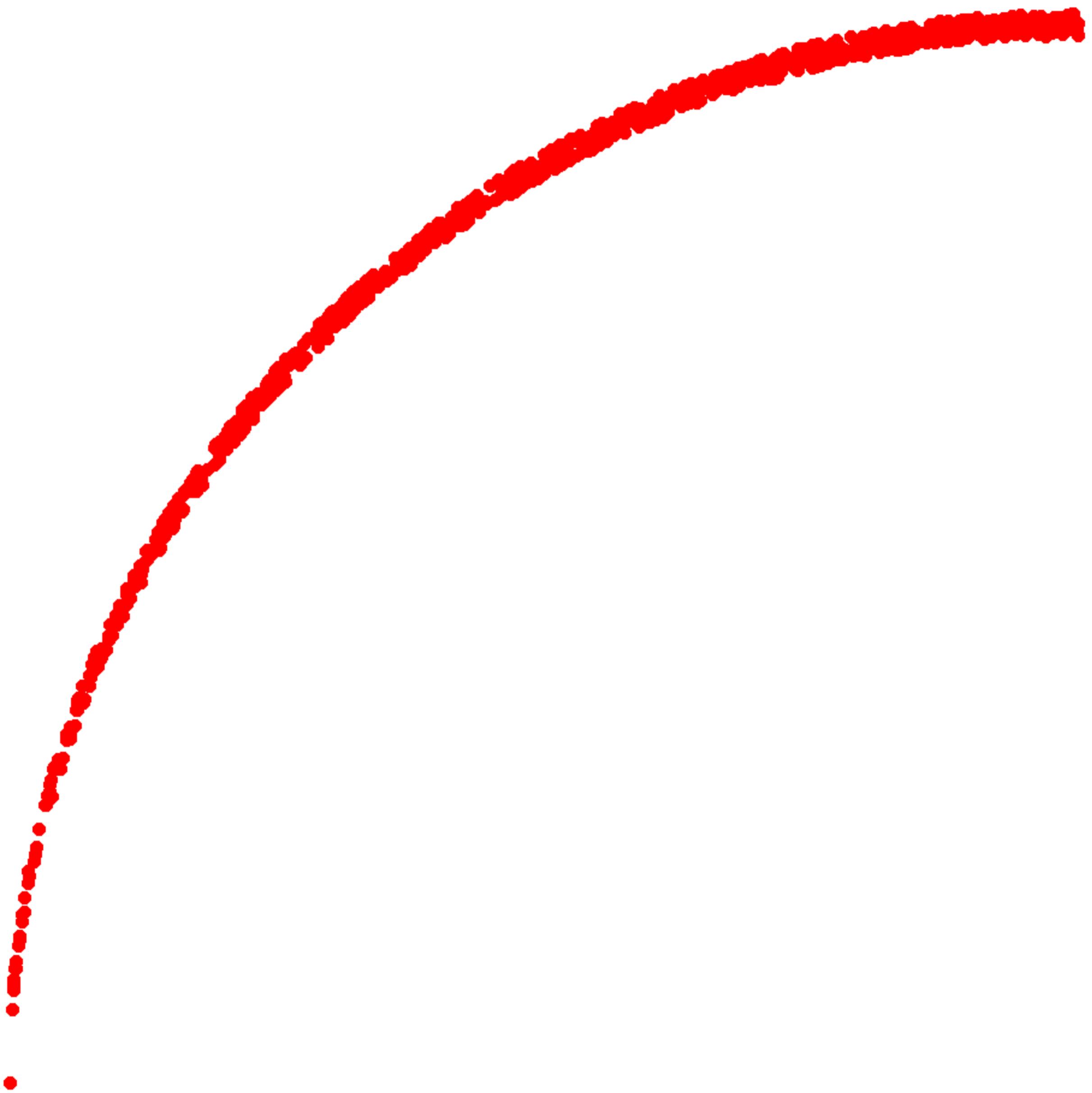
0.919



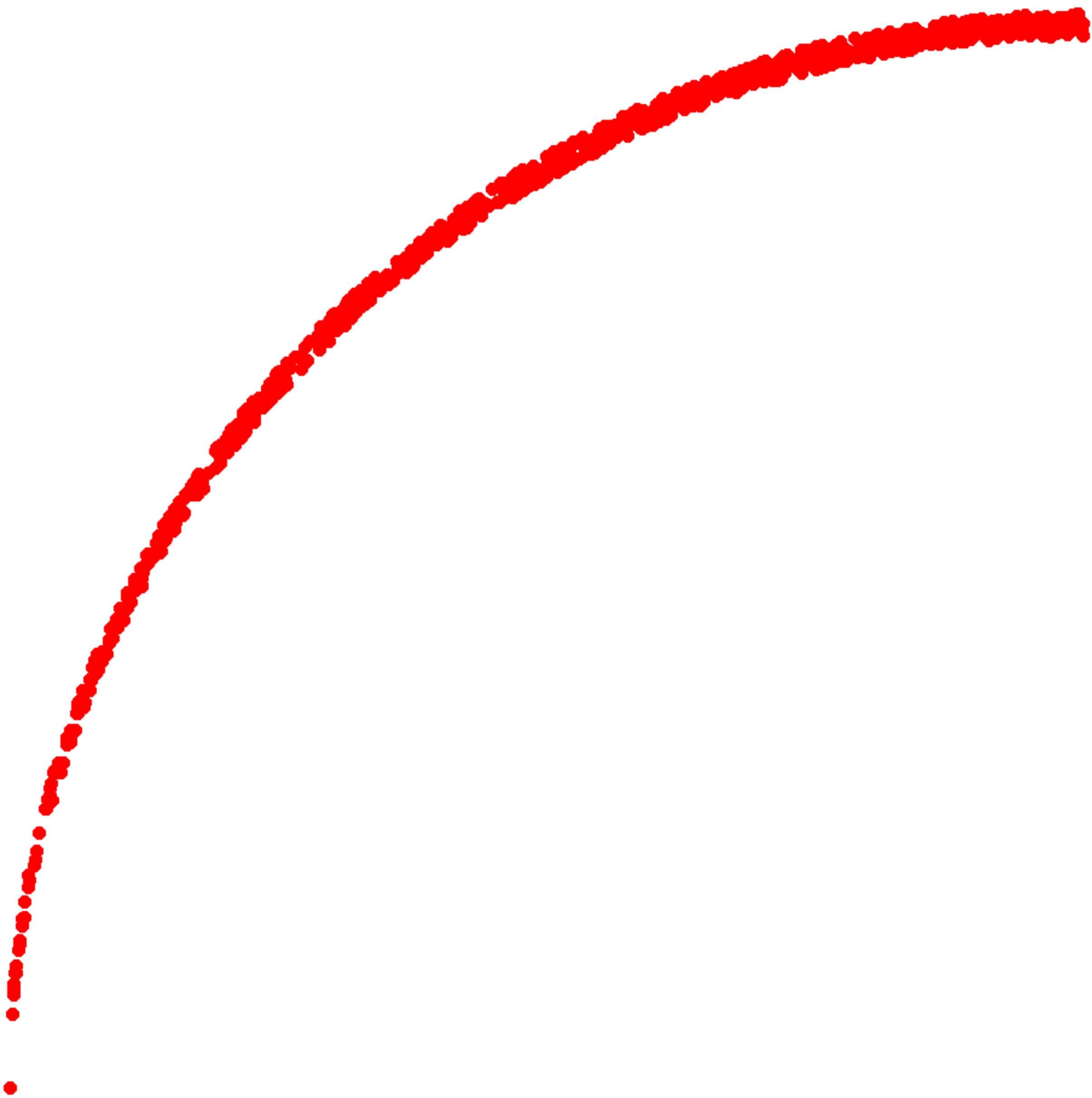
0.92005



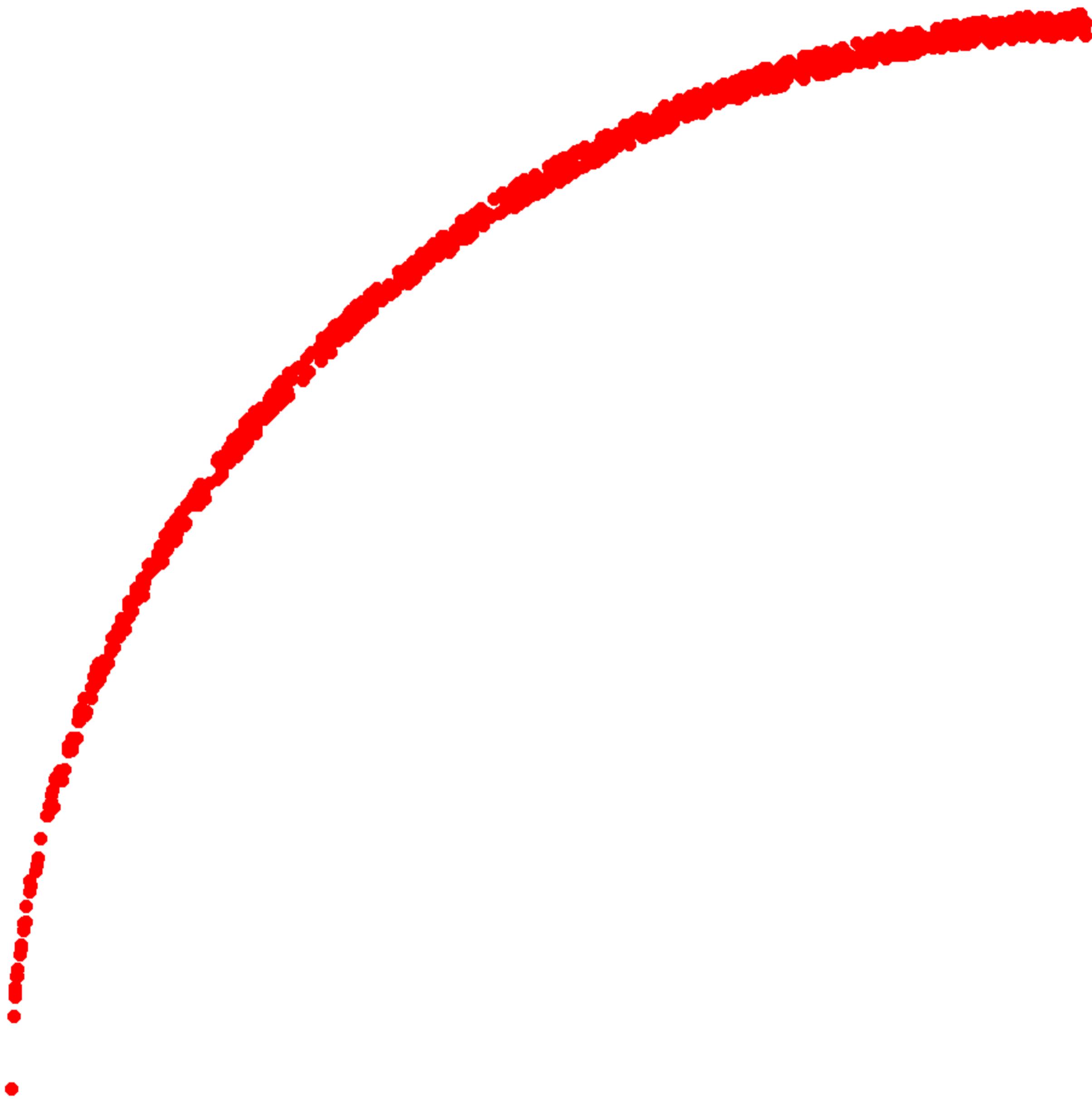
0.92104



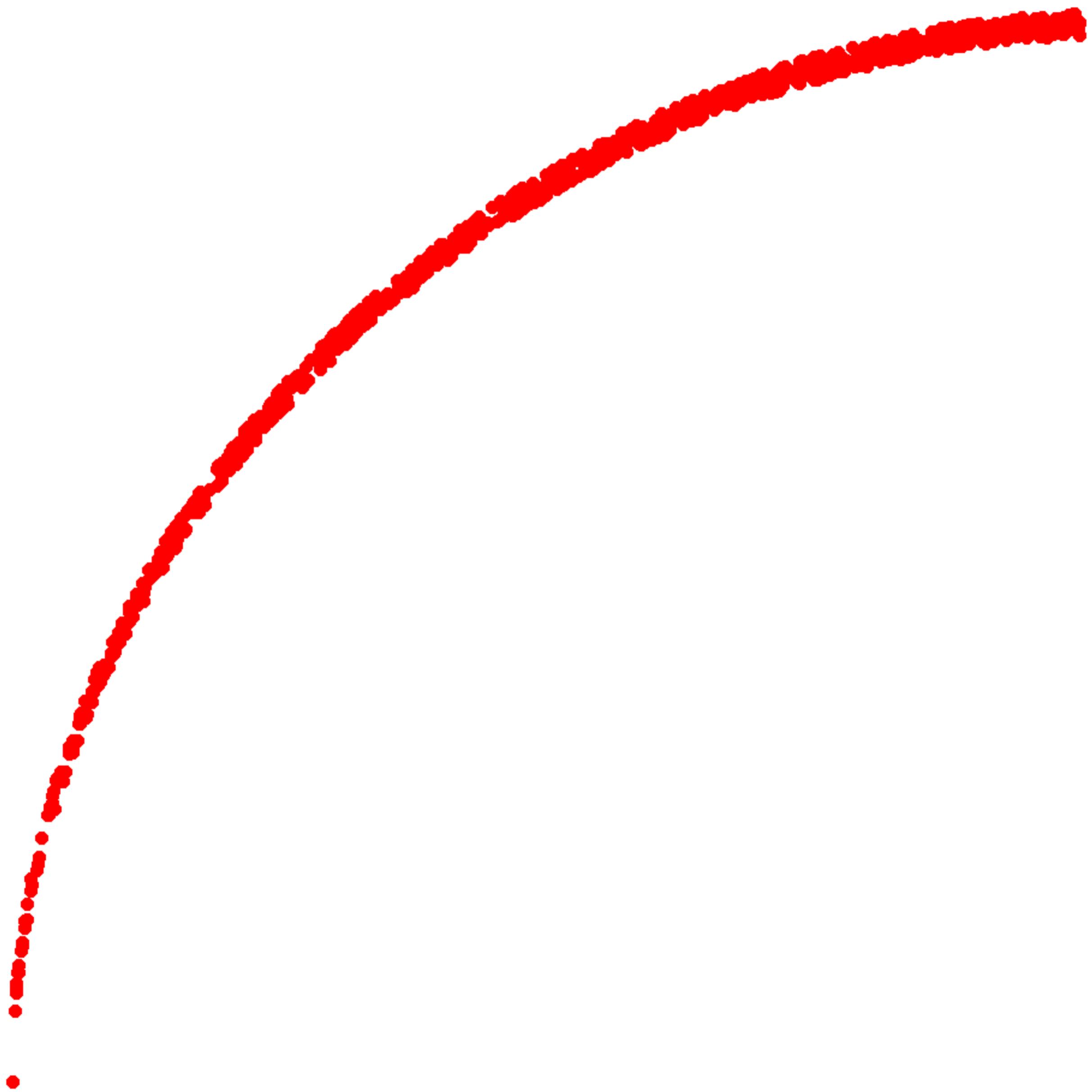
0.92197



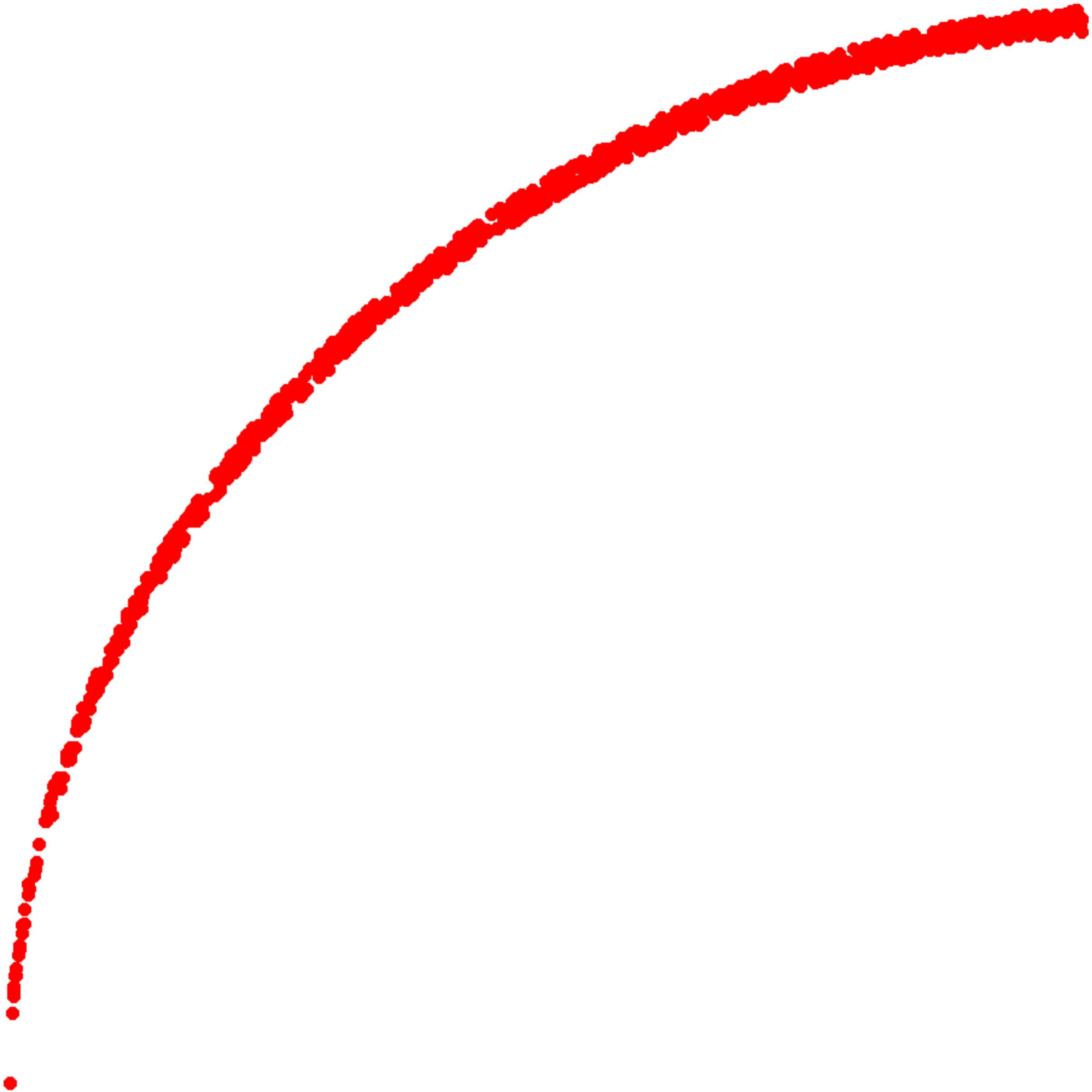
0.92675



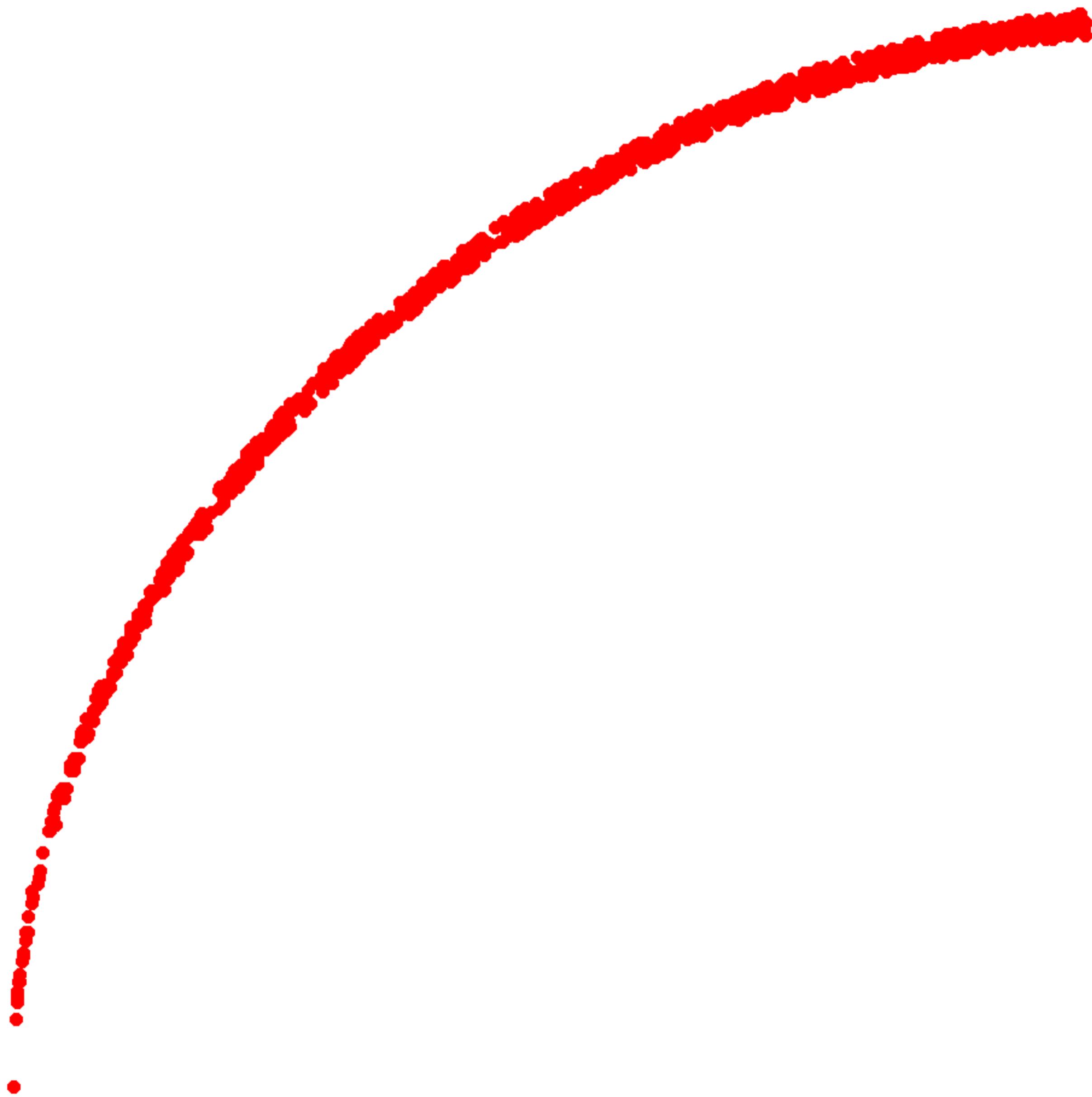
0.93134



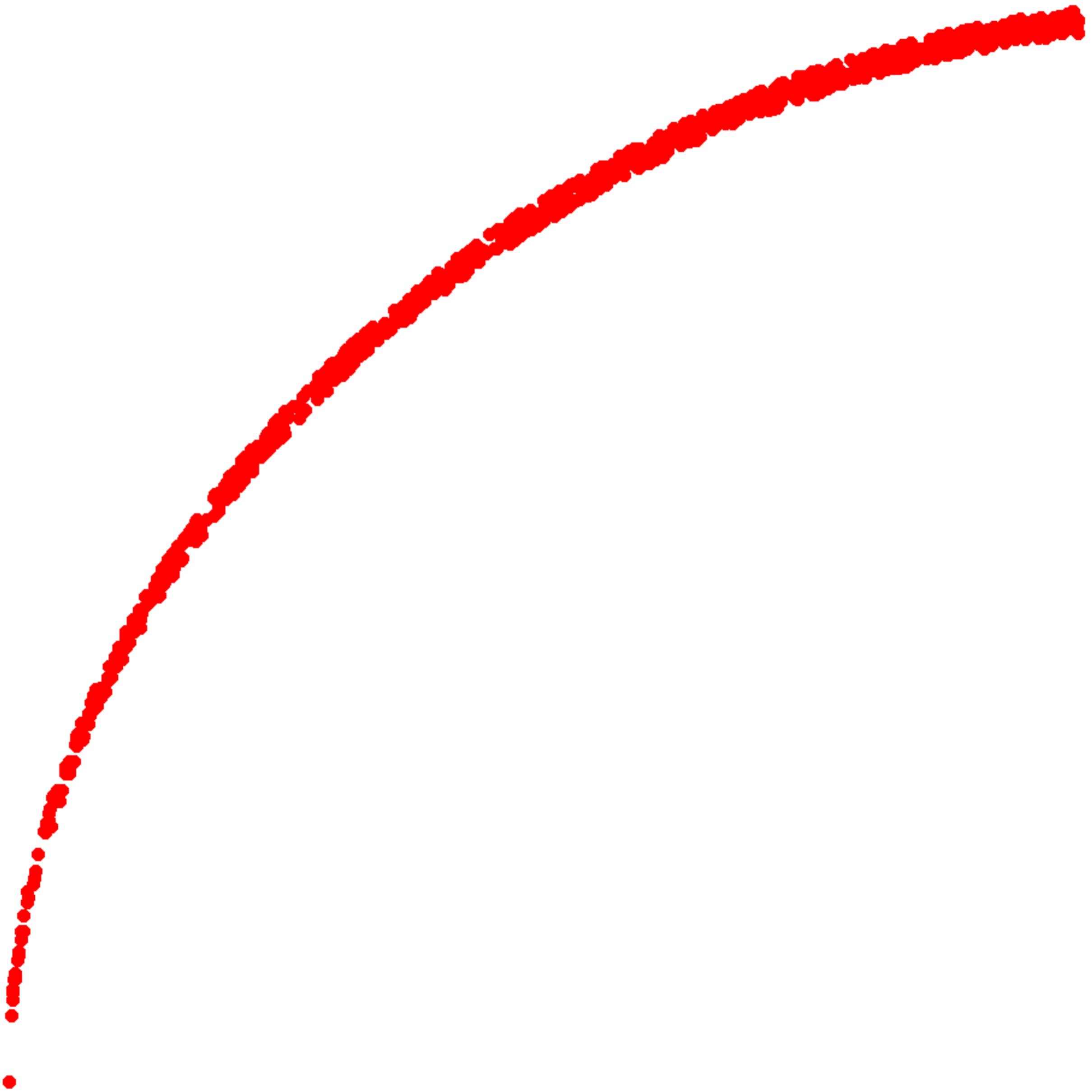
0.93573



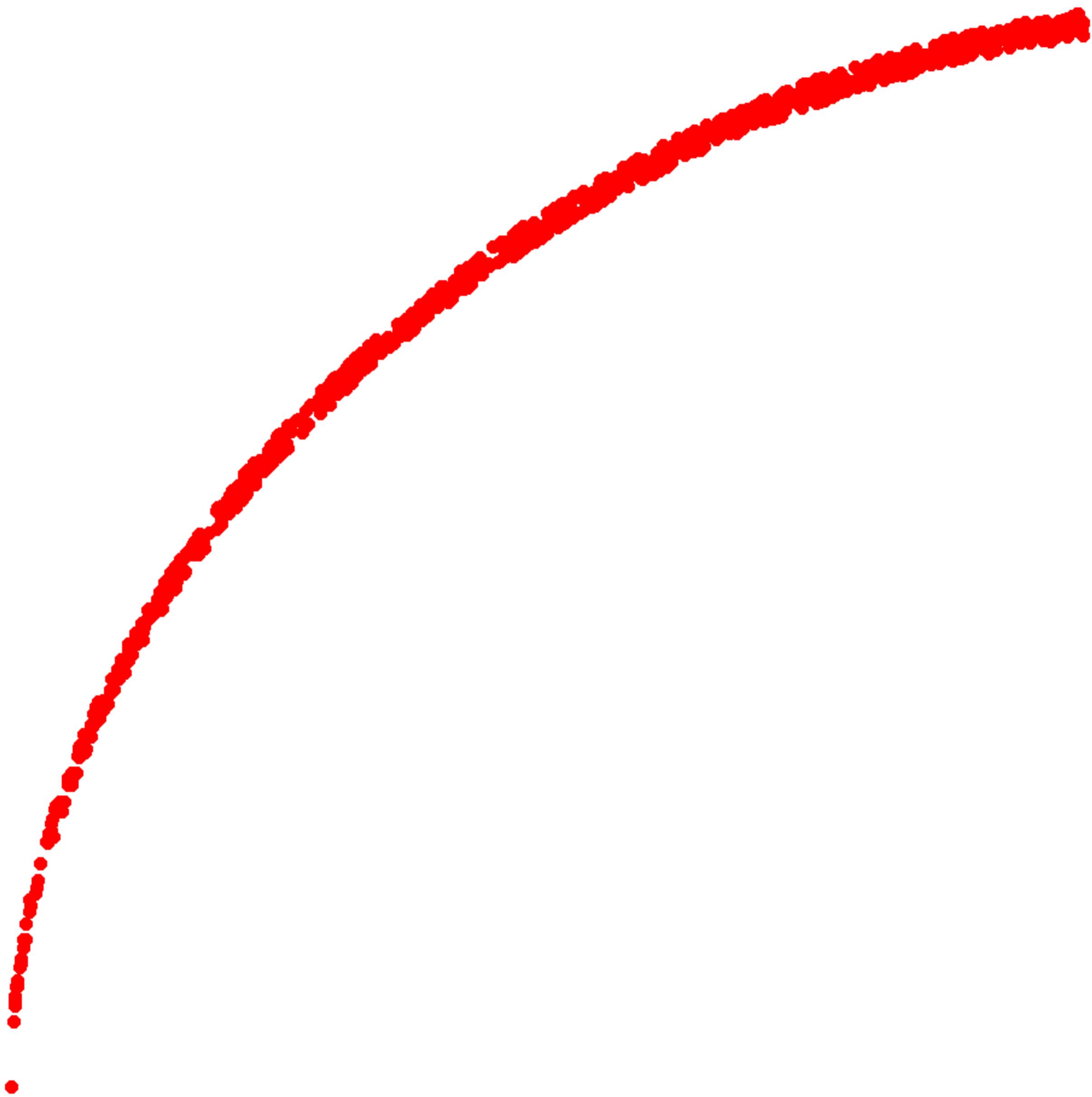
0.93993



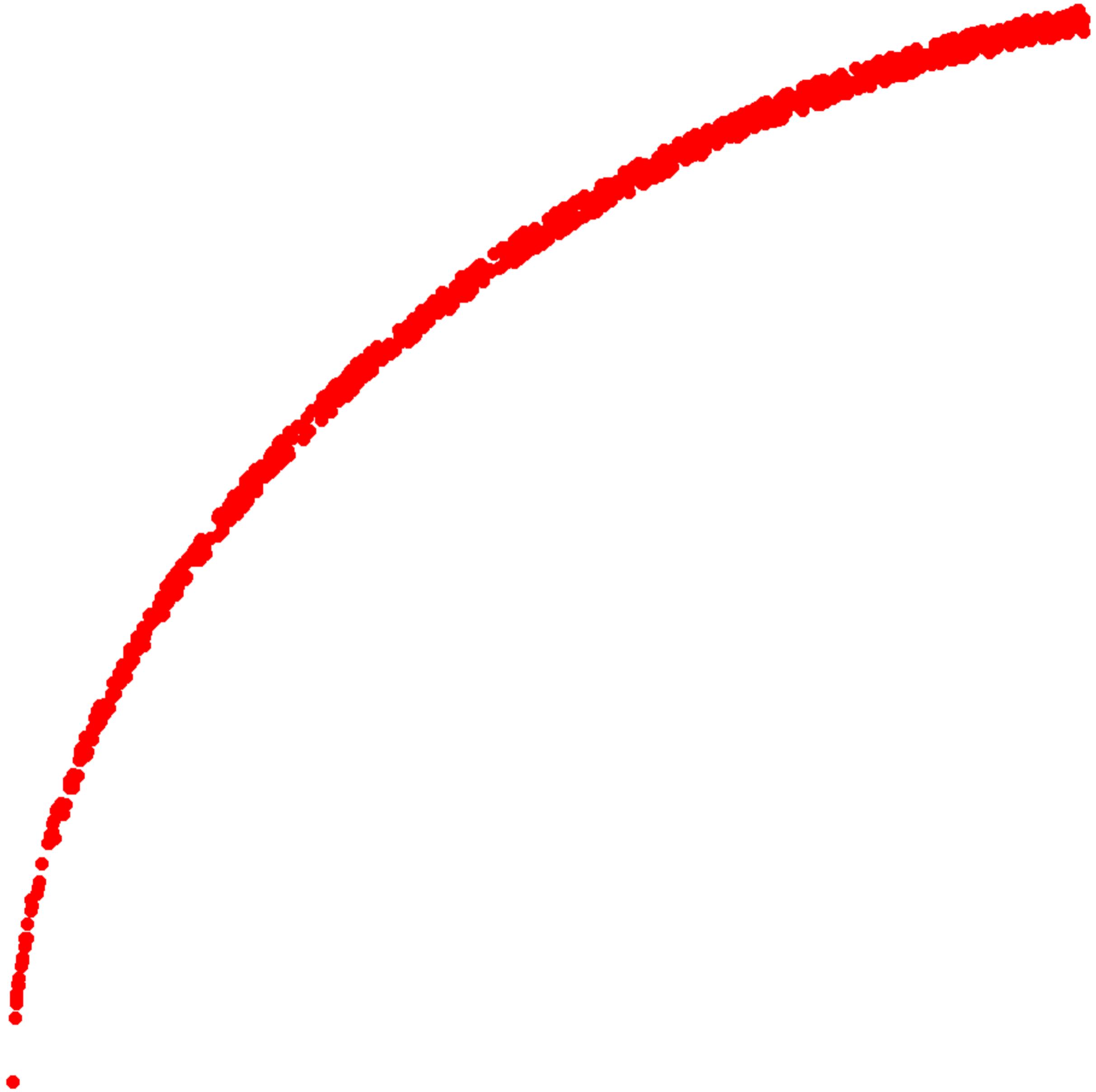
0.94396



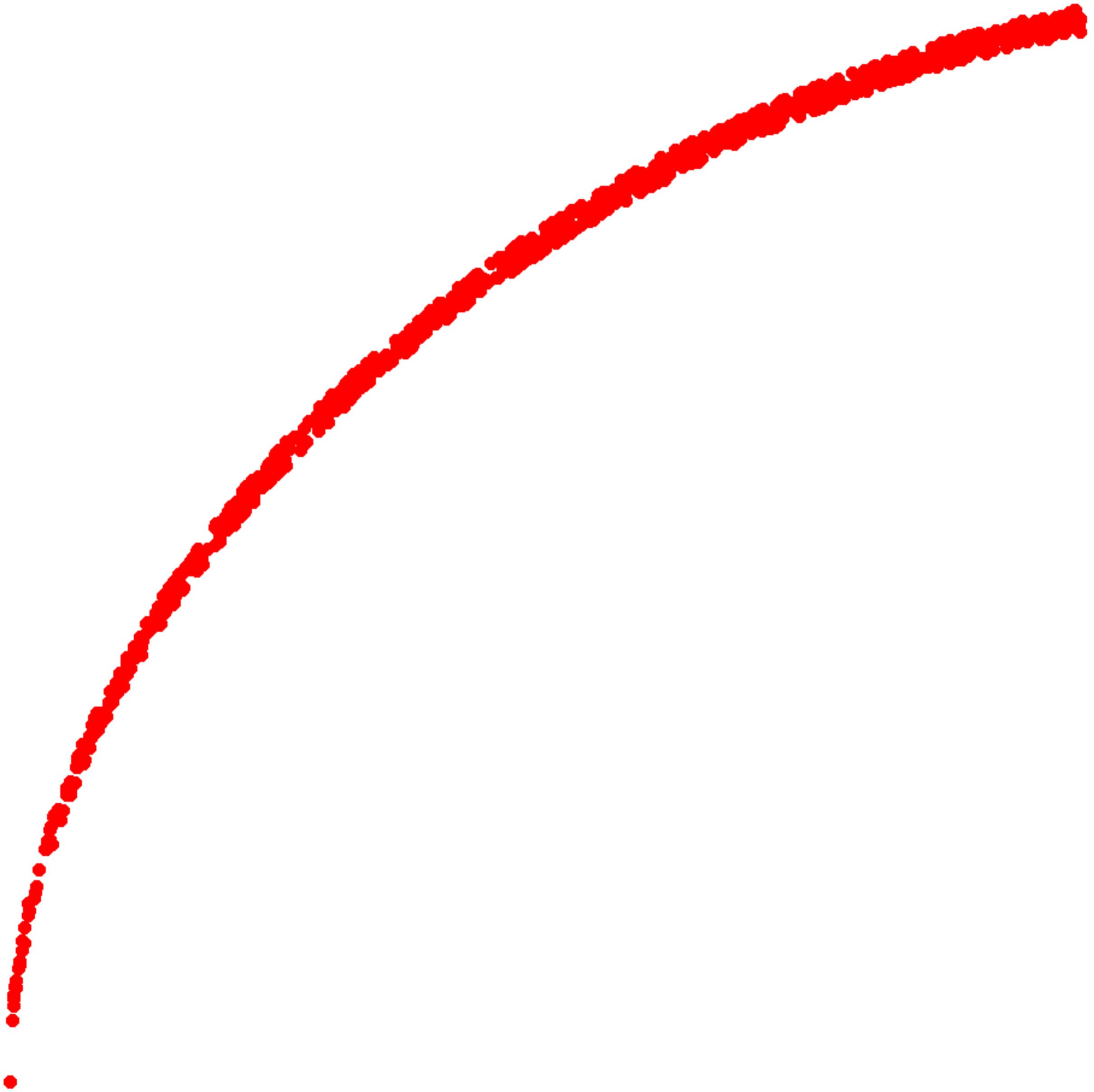
0.94779



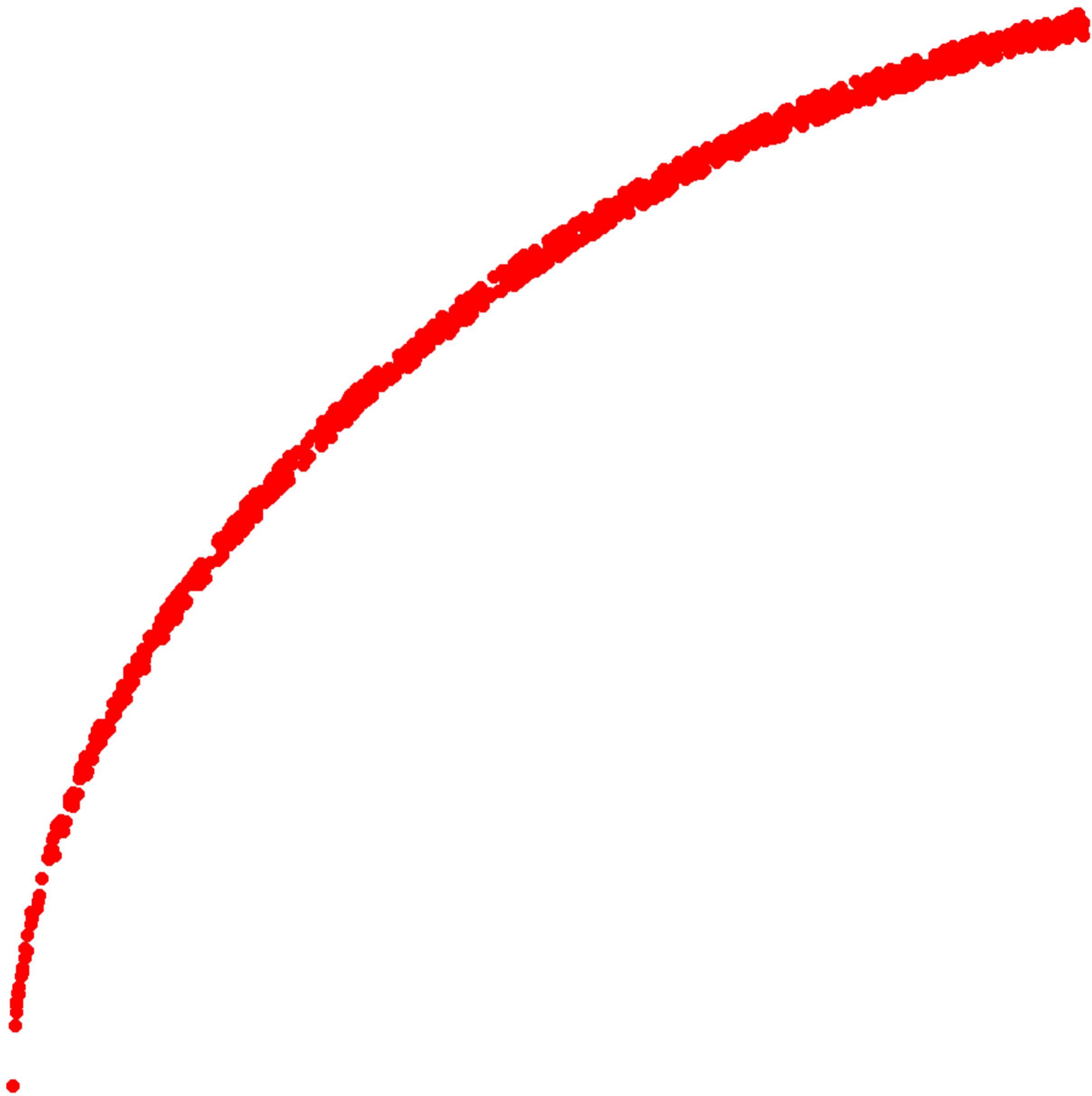
0.95146



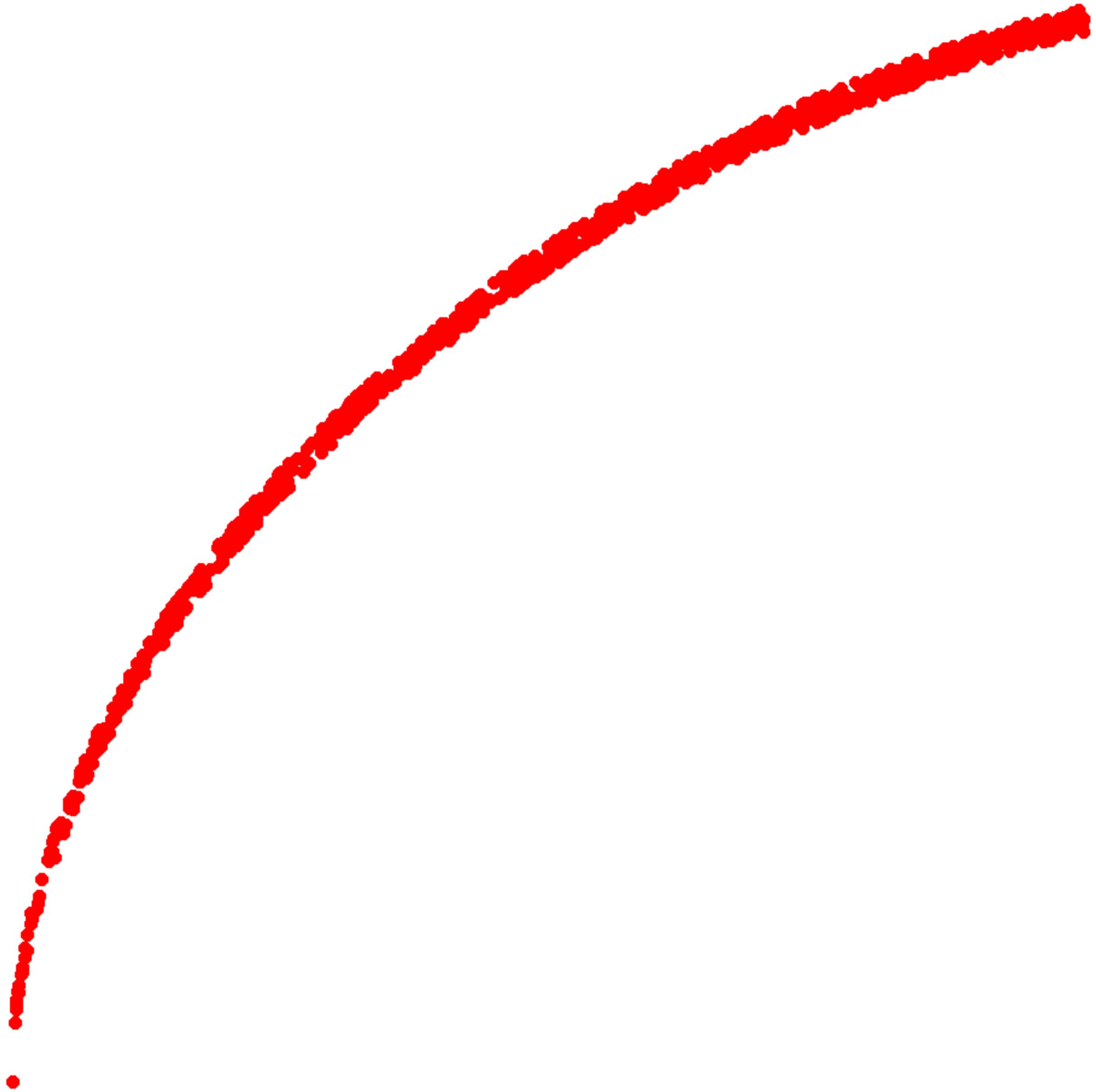
0.95495



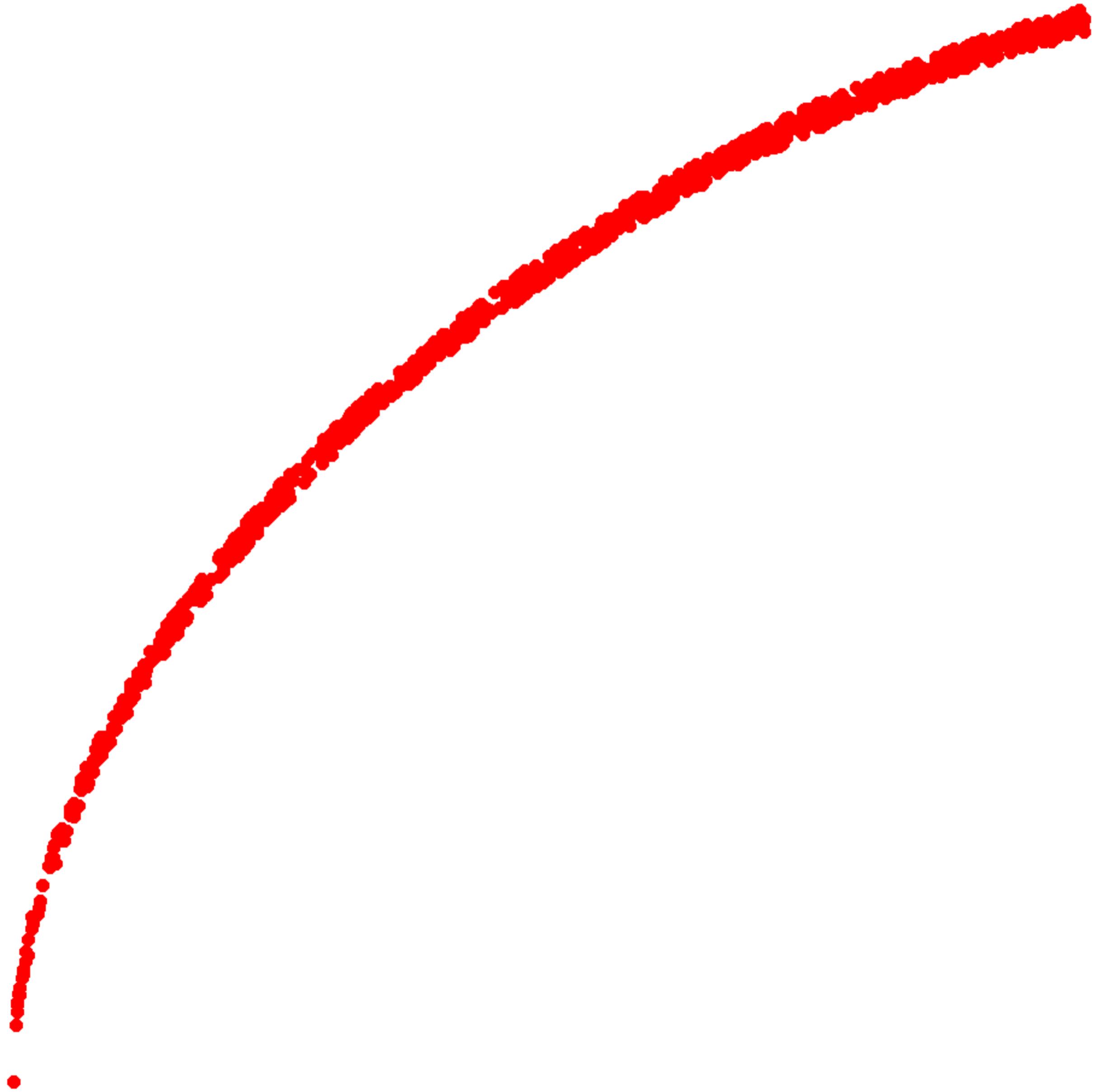
0.95827



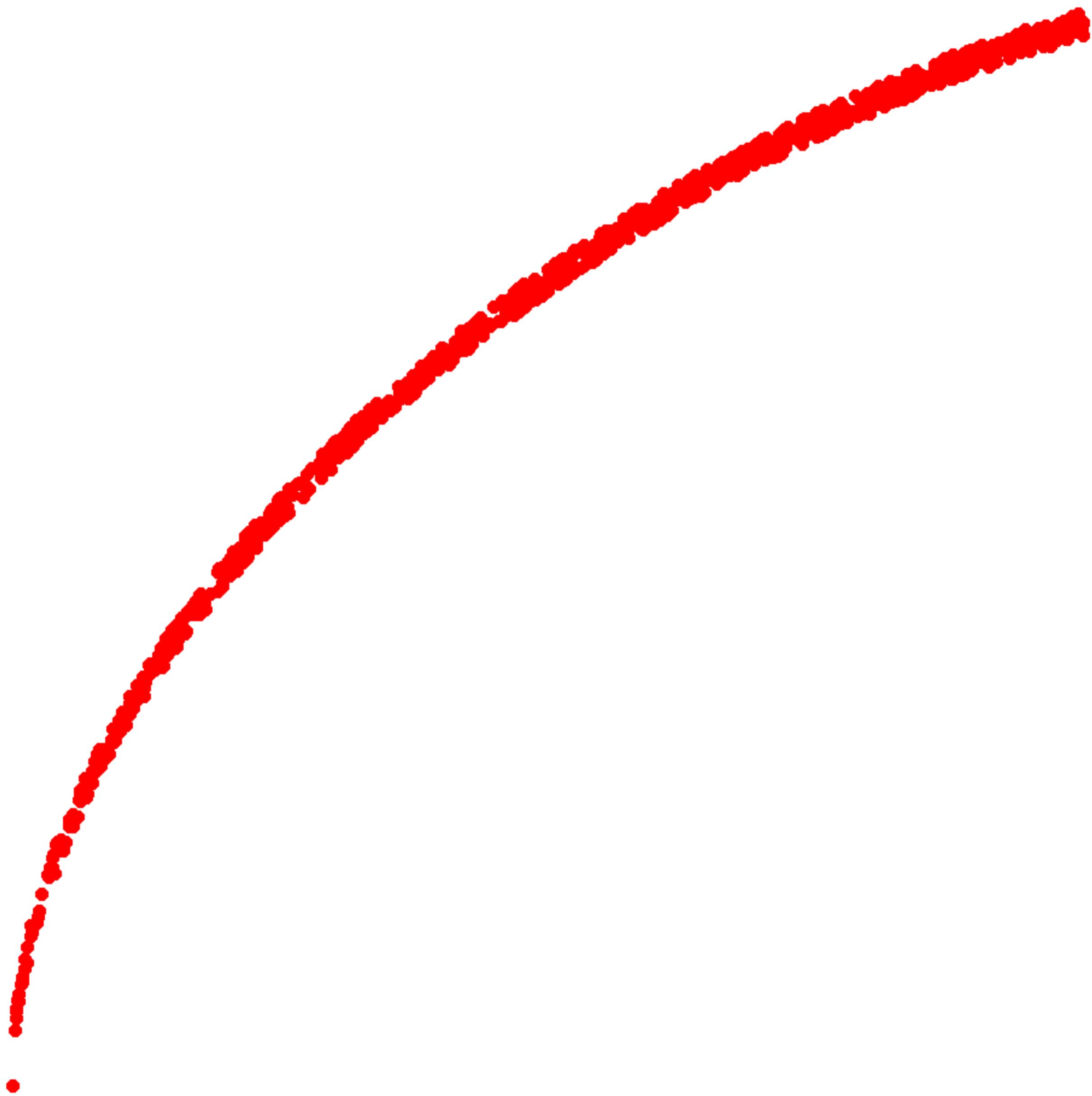
0.96143



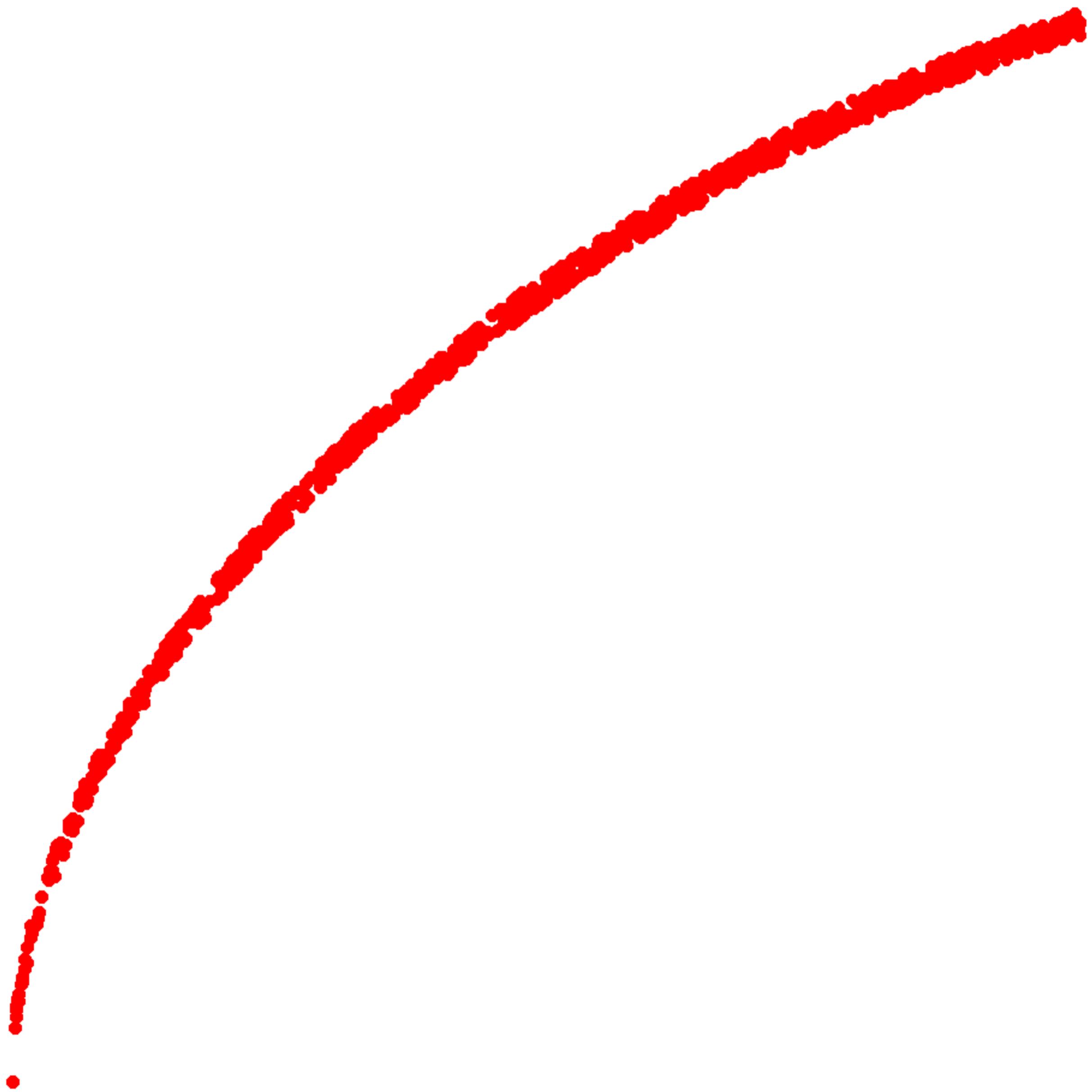
0.96442



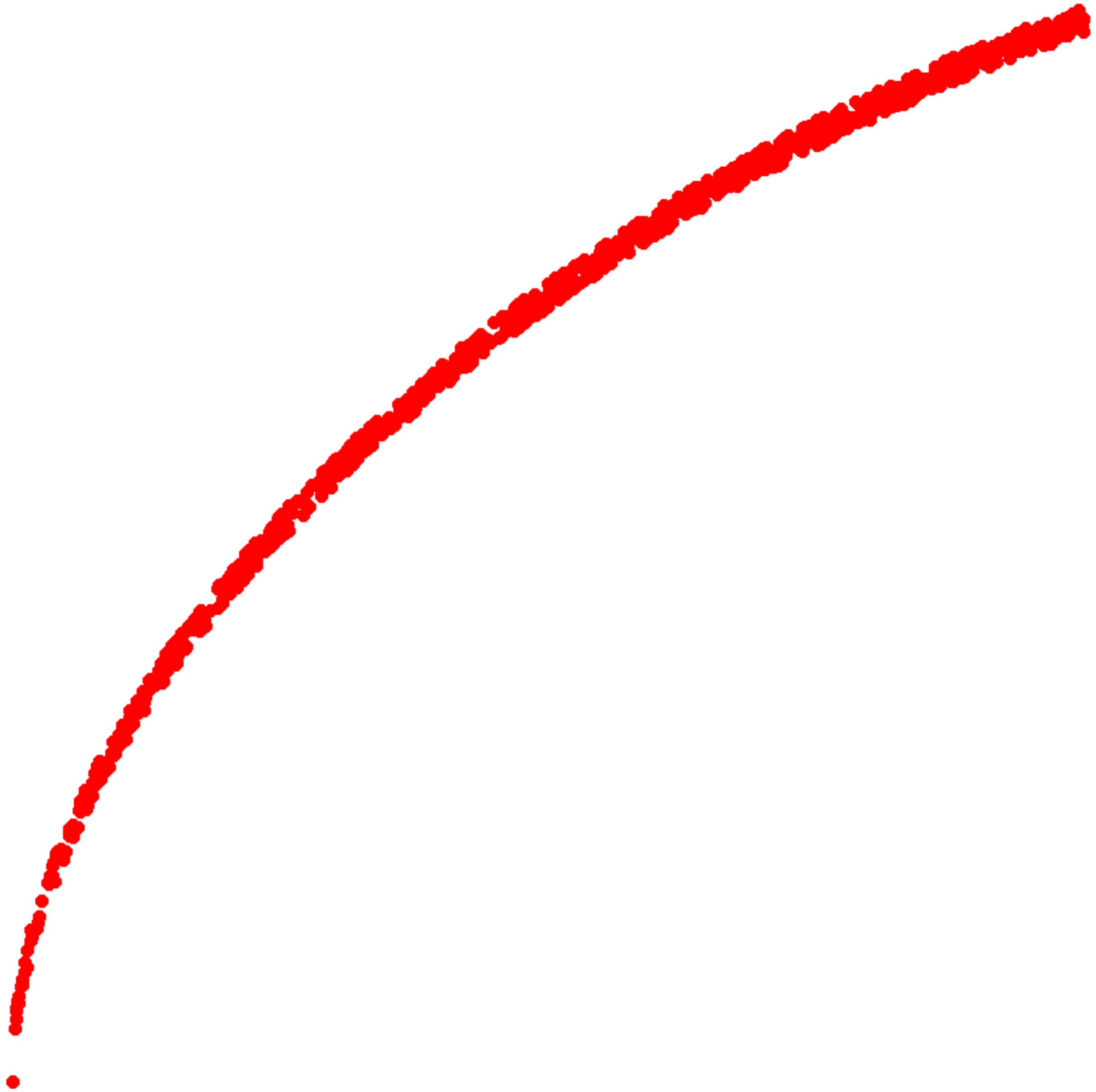
0.96726



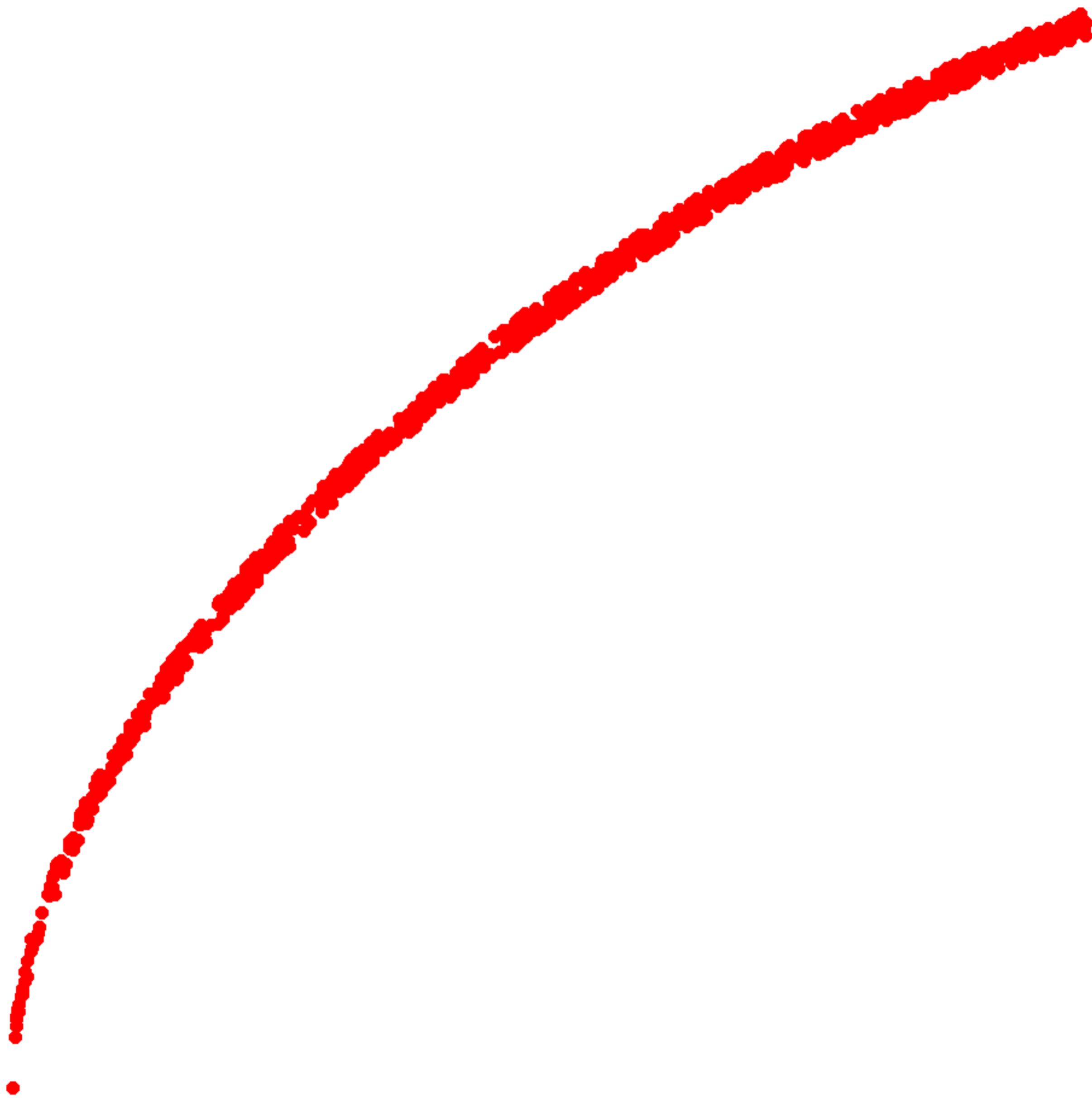
0.96995



0.97249



0.97489



0.97715



0.97927



0.98126



0.98313



0.98487



0.98649



0.988



0.98939



0.99068



0.99187



0.99296



0.99395



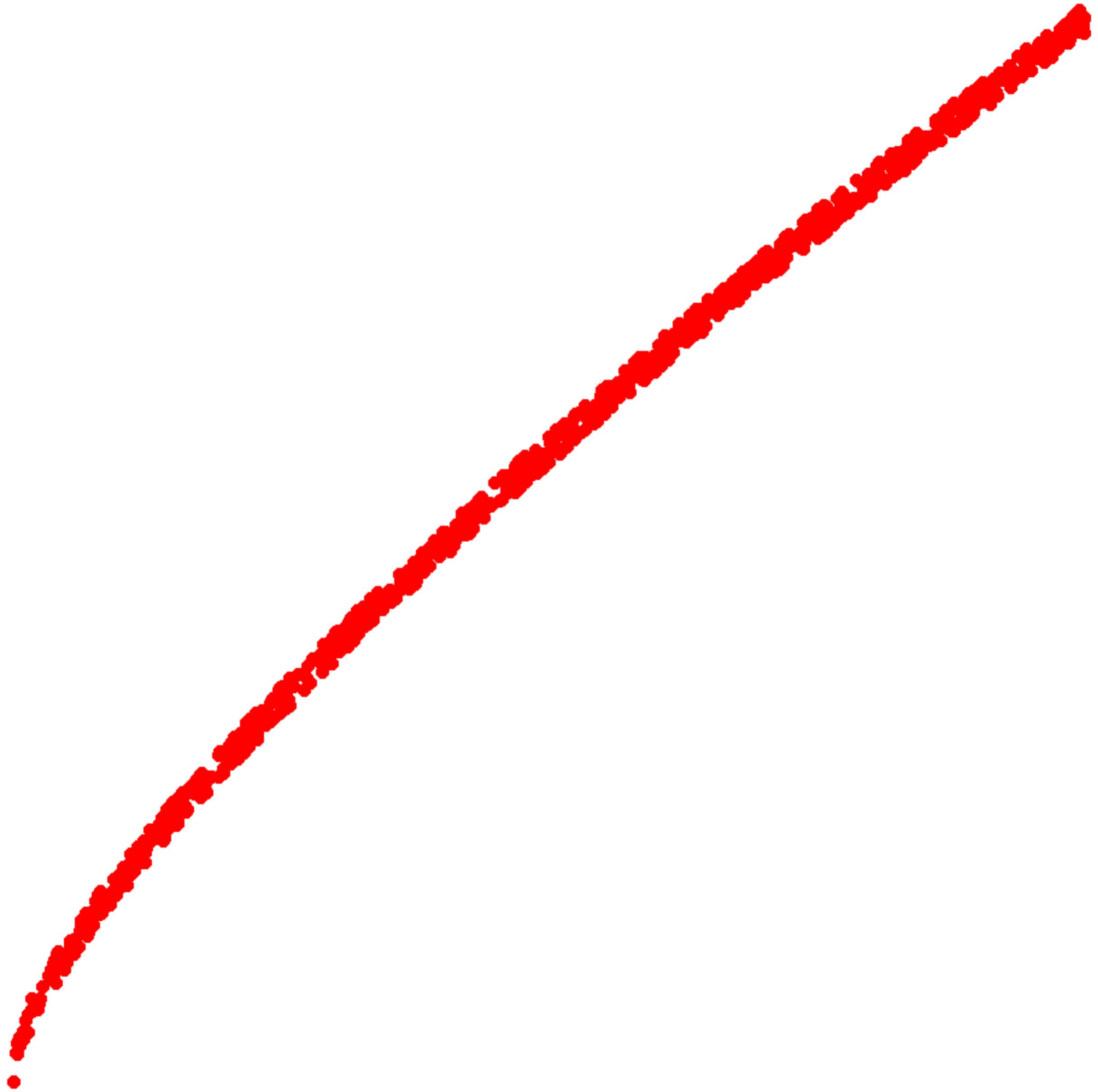
0.99485



0.99565



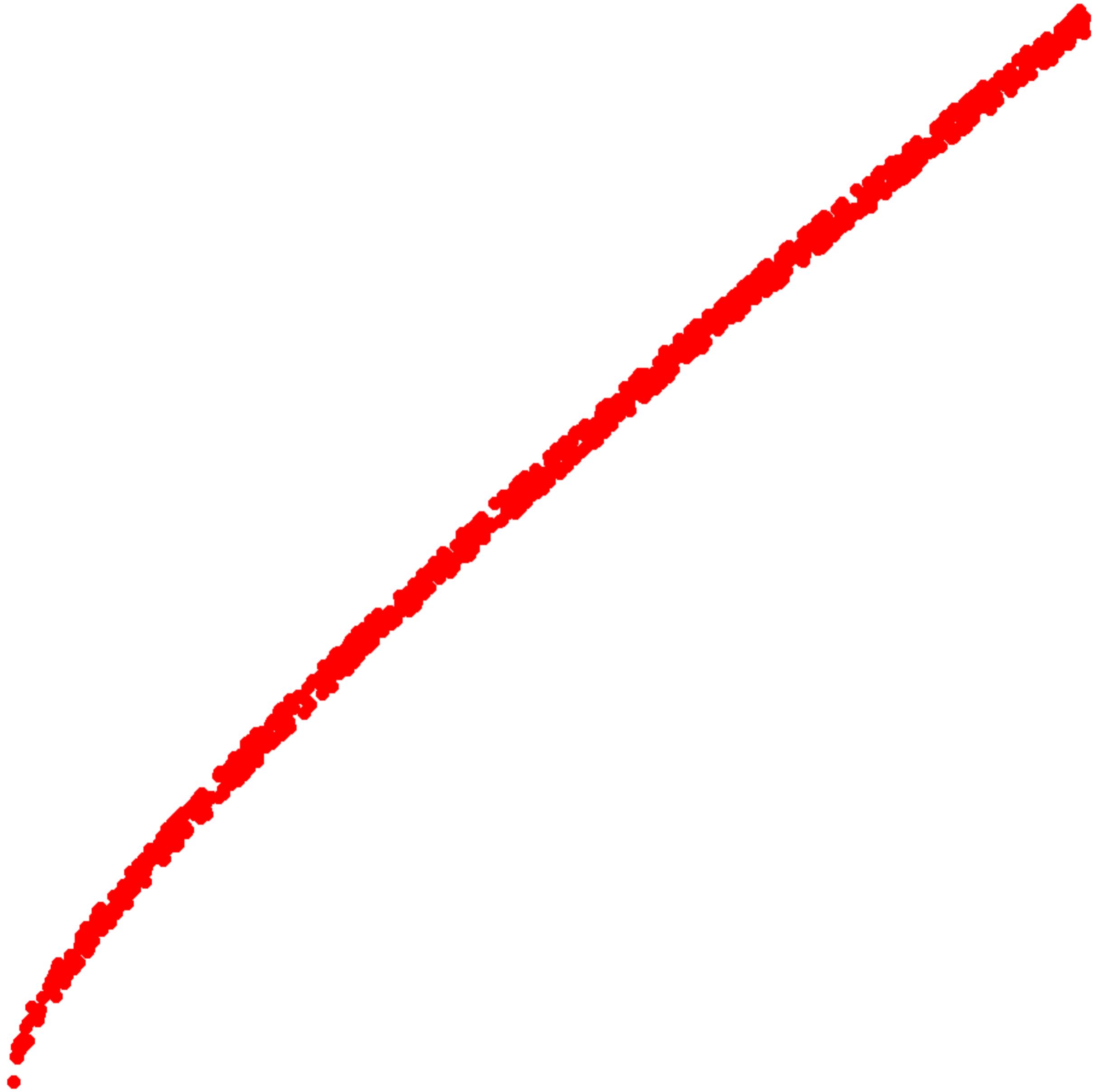
0.99638



0.99702



0.99758



0.99807



0.99848



0.99882



0.9991



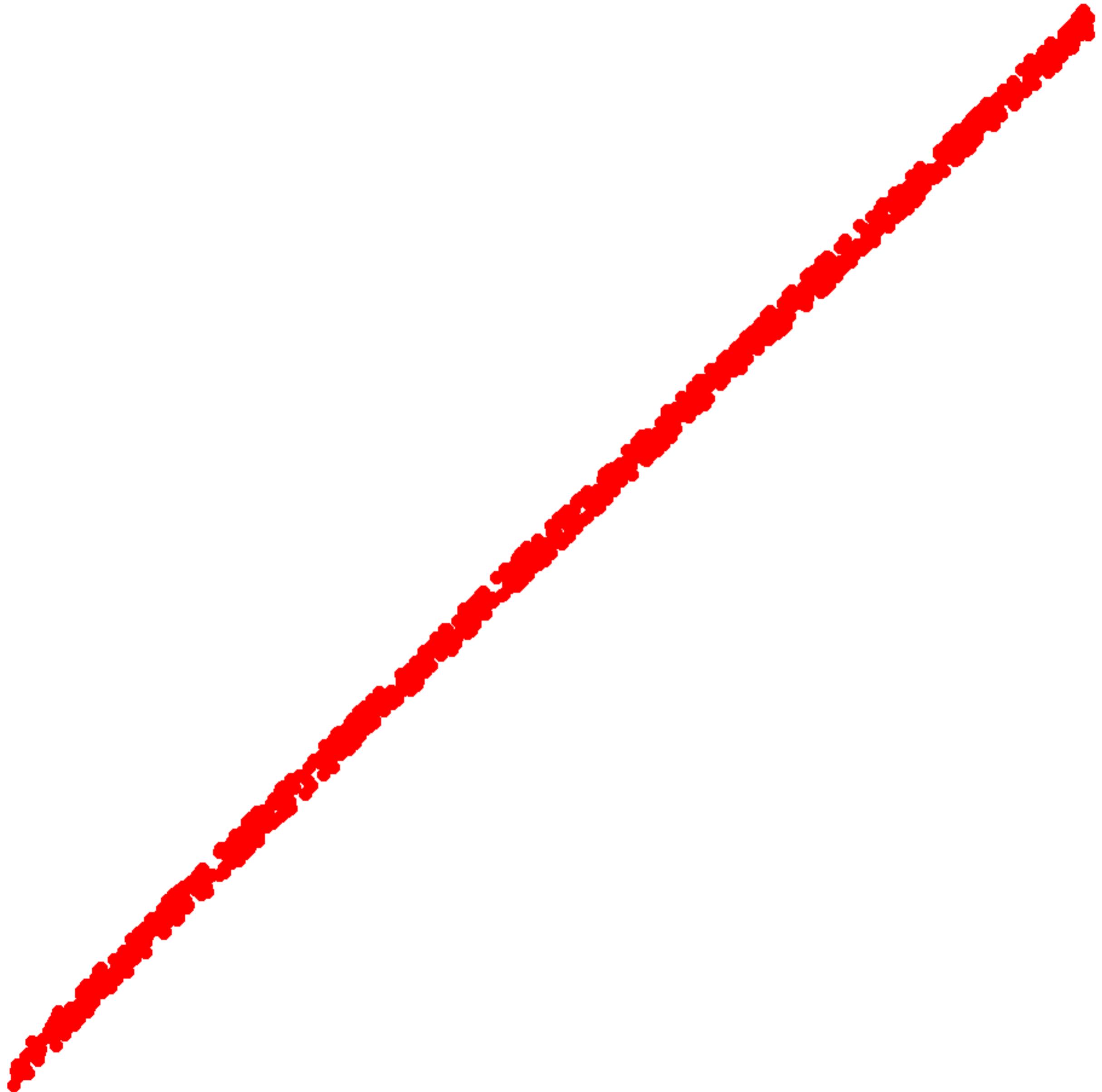
0.99931



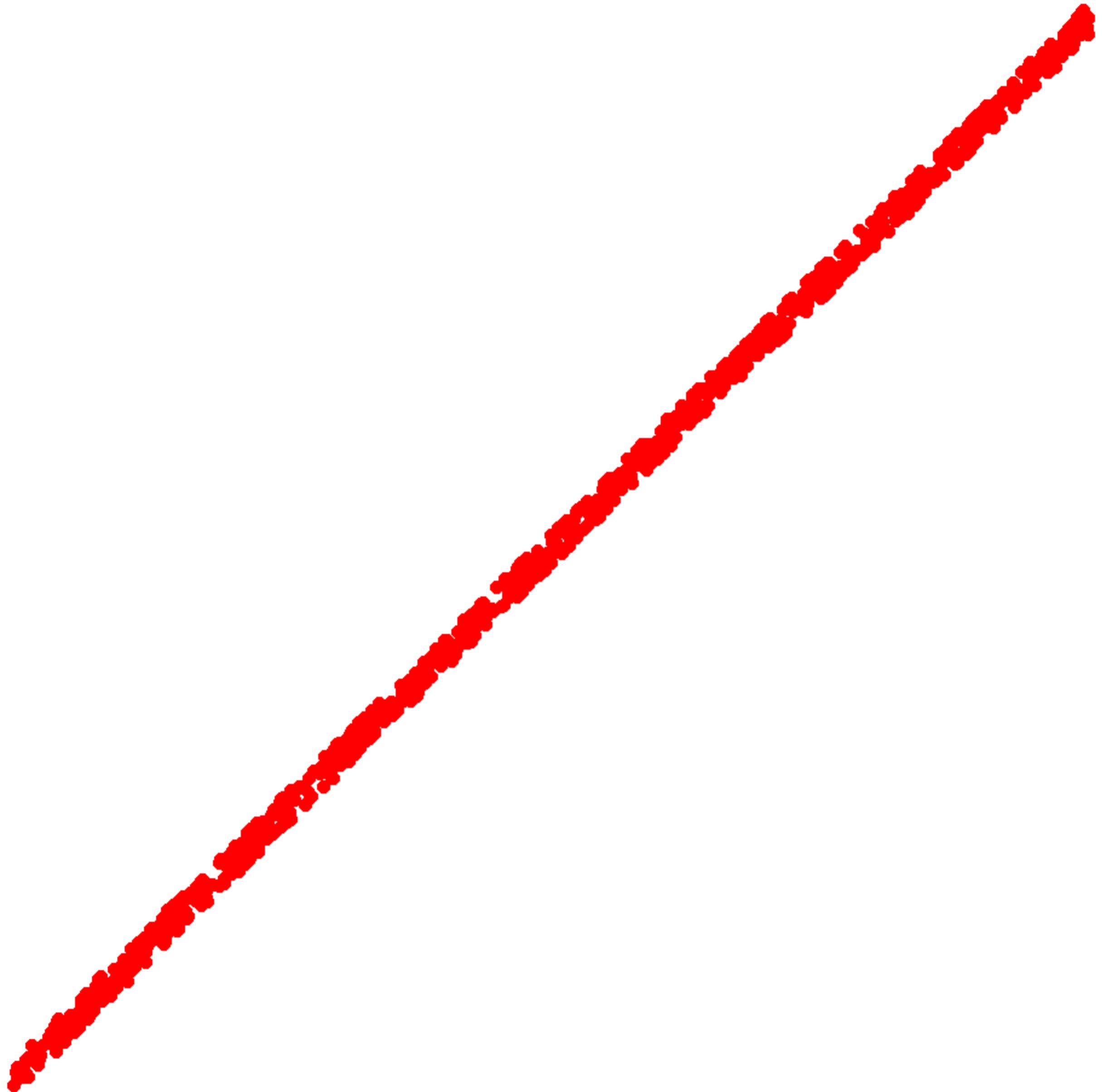
0.99946



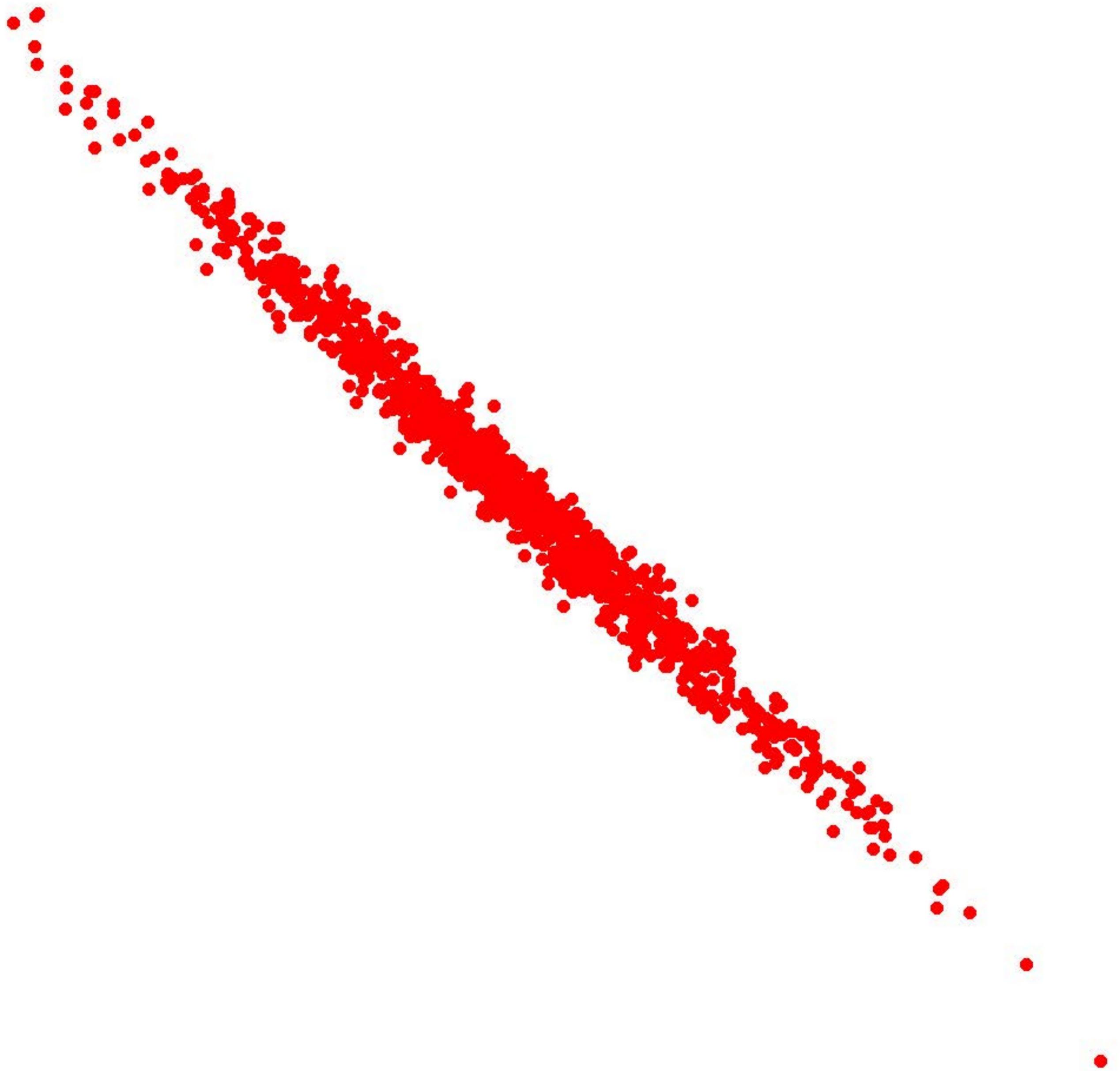
0.99955



0.99958



-0.99



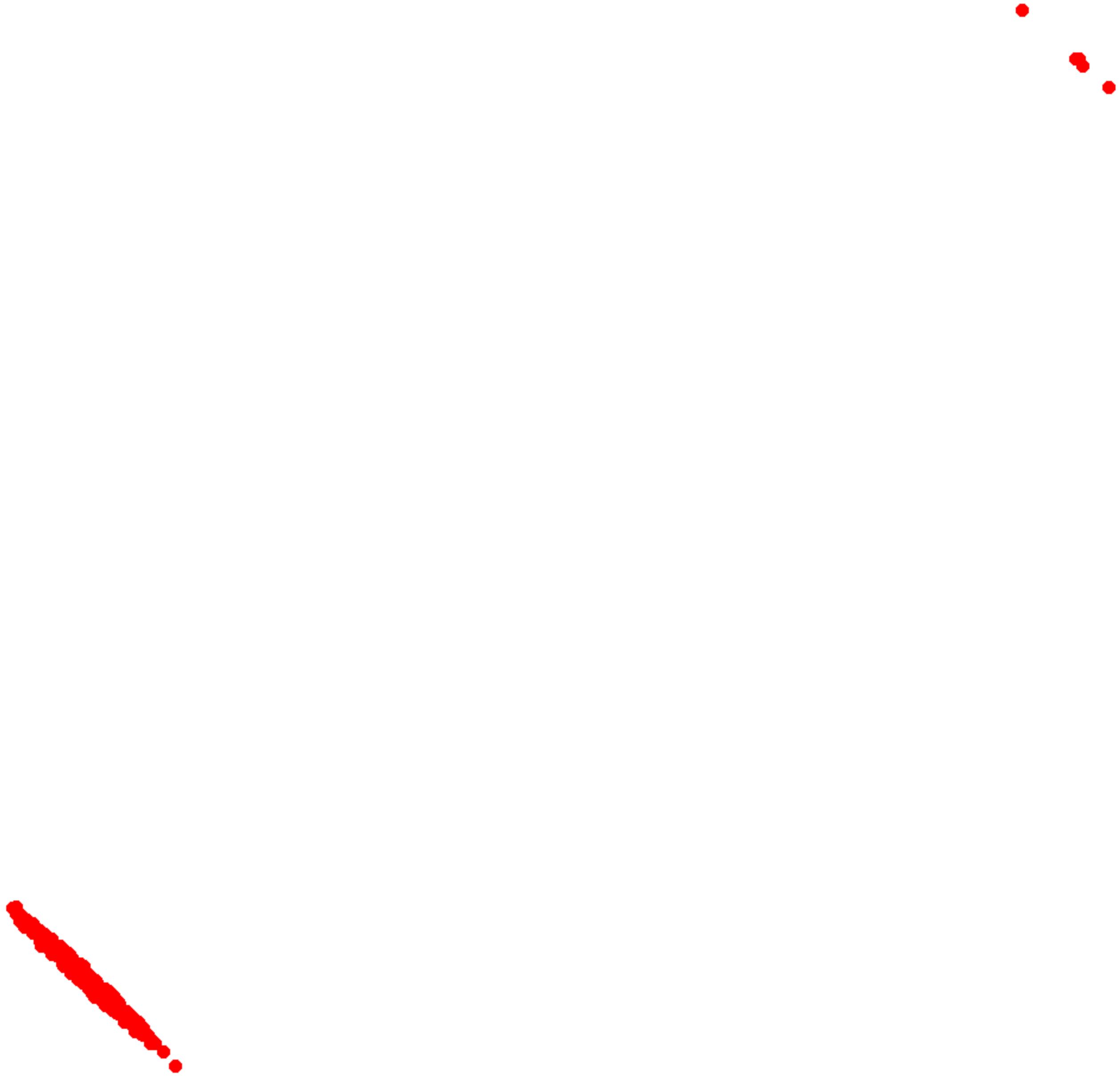
0.78978



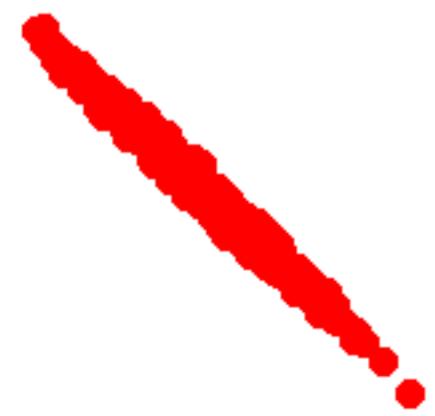
0.78018



0.77008



0.75943



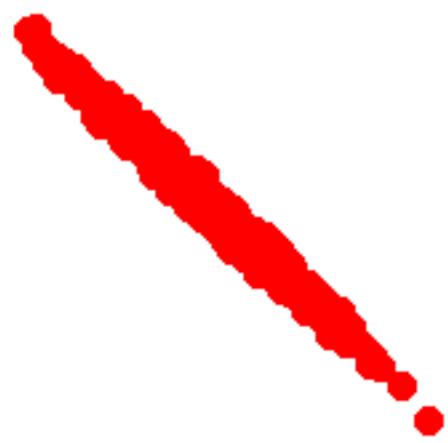
0.74822



0.7364



0.72395



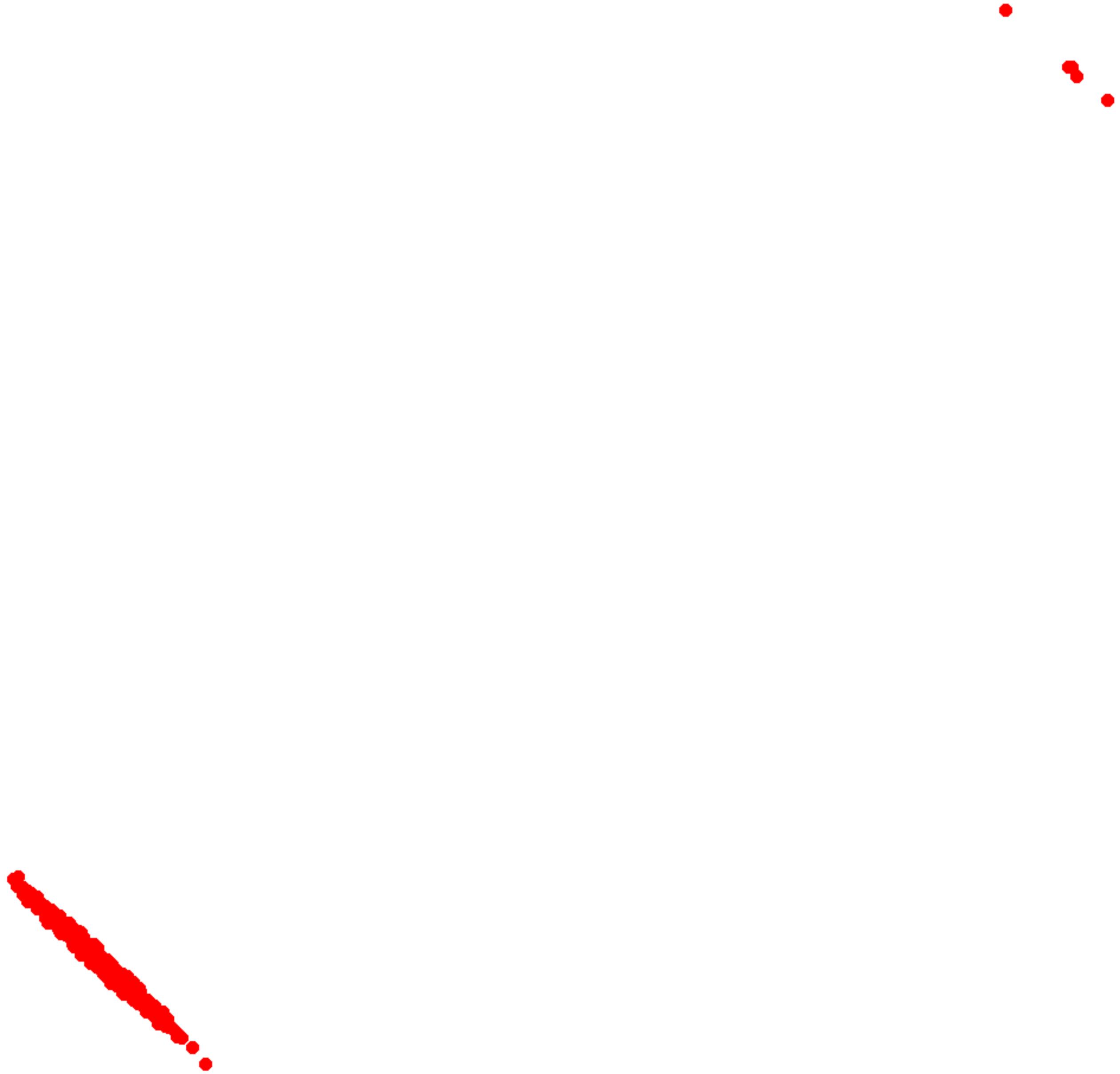
0.71083



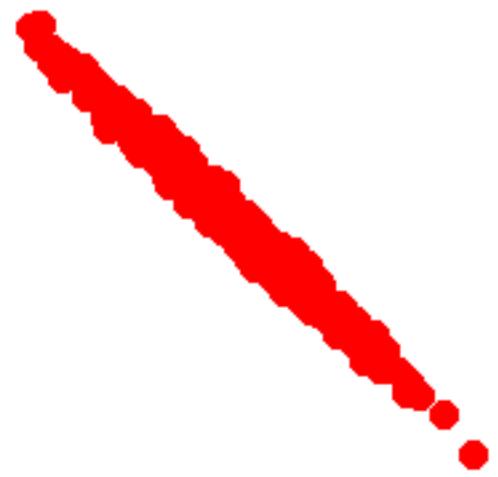
0.69701



0.68245



0.6671



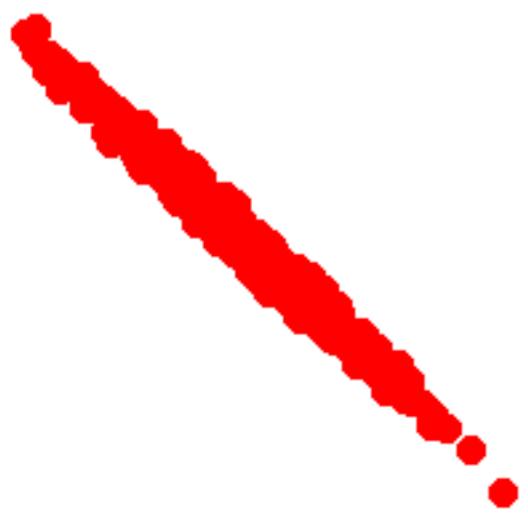
0.65094



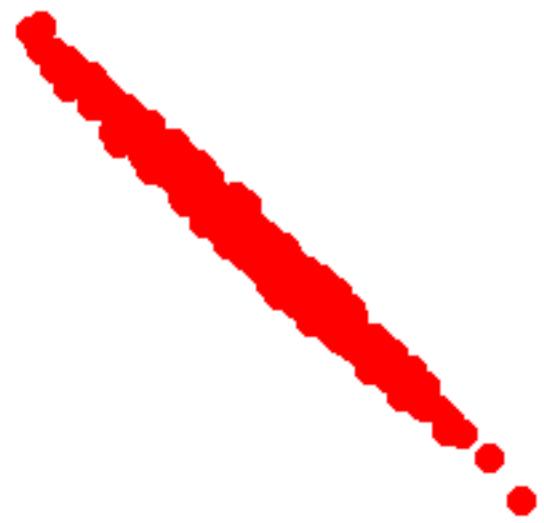
0.63392



0.616



0.59713



0.57728



0.5564



0.53445



0.51139



0.48718



0.46178



0.43516



0.40728



0.37813



0.34768



0.31592



0.28283



0.24844



0.21274



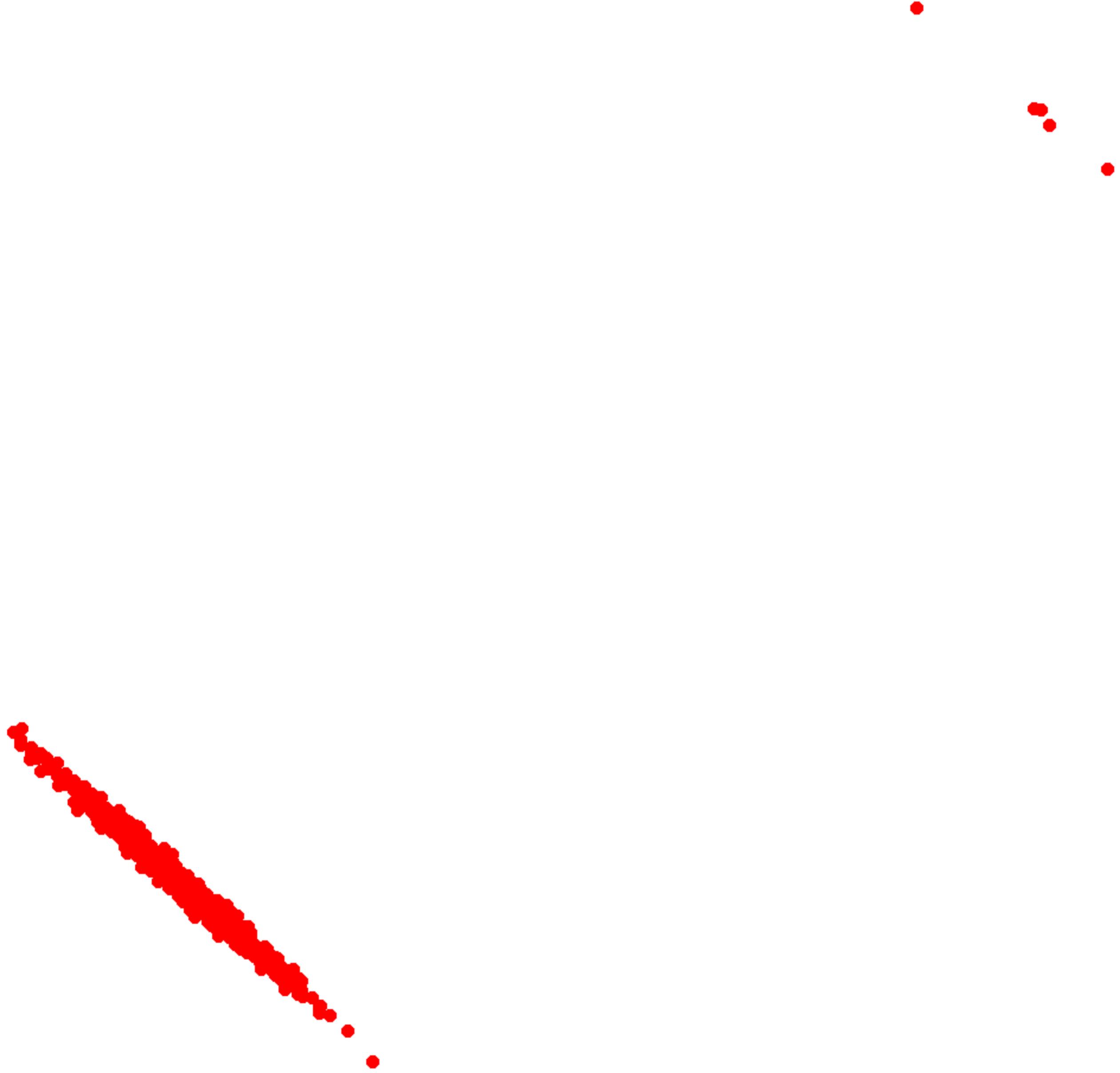
0.17577



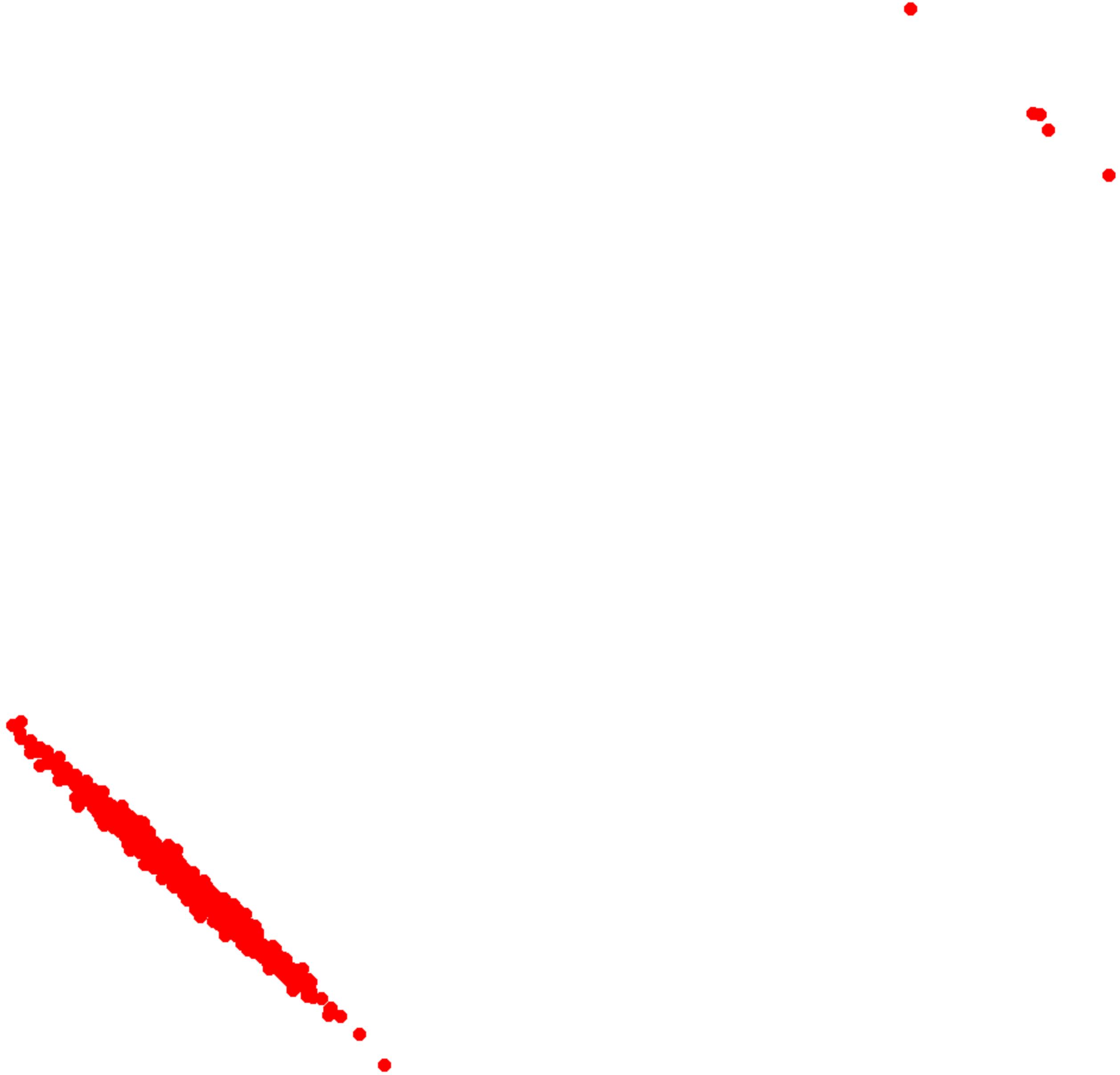
0.13756



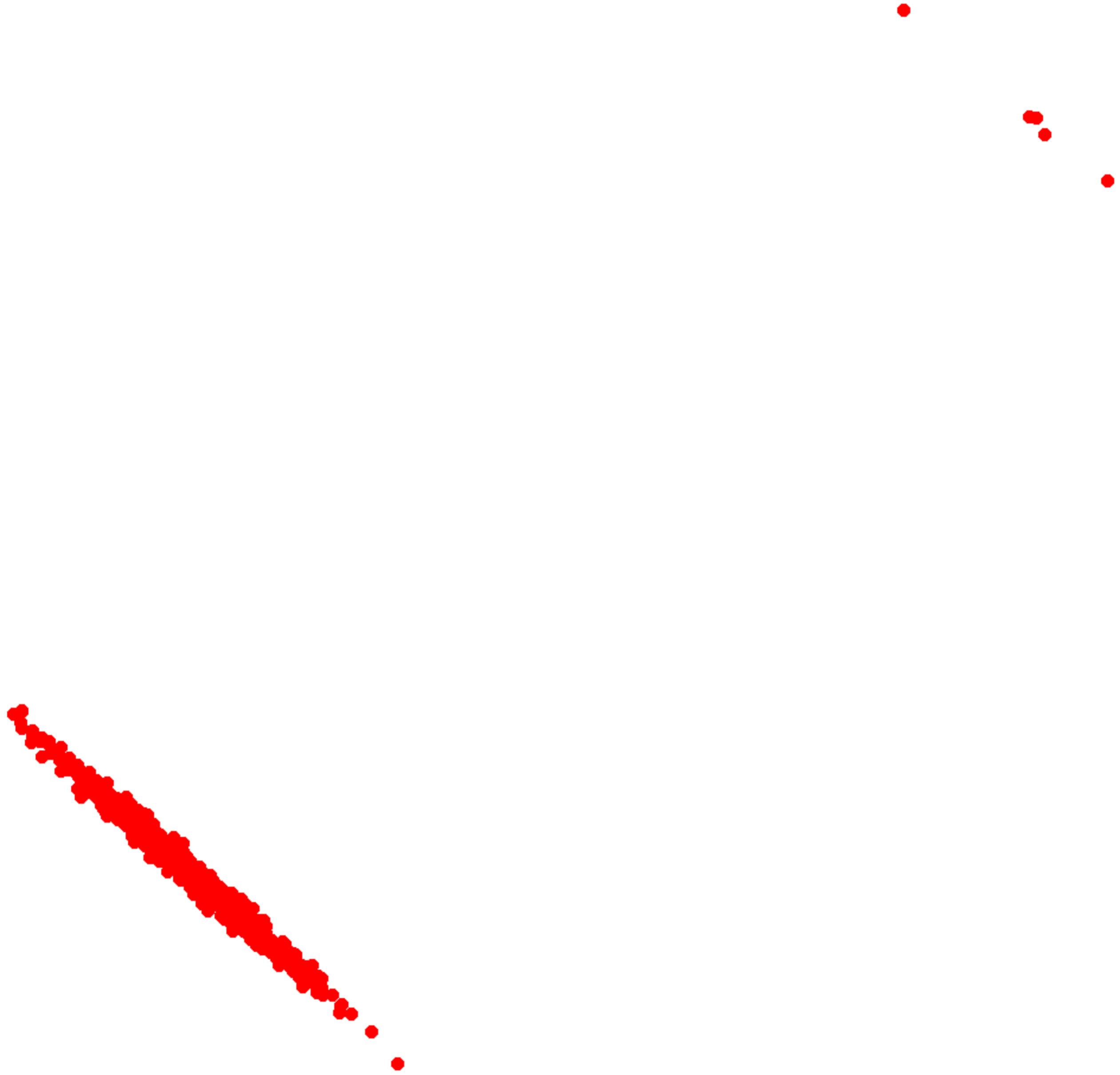
0.09817



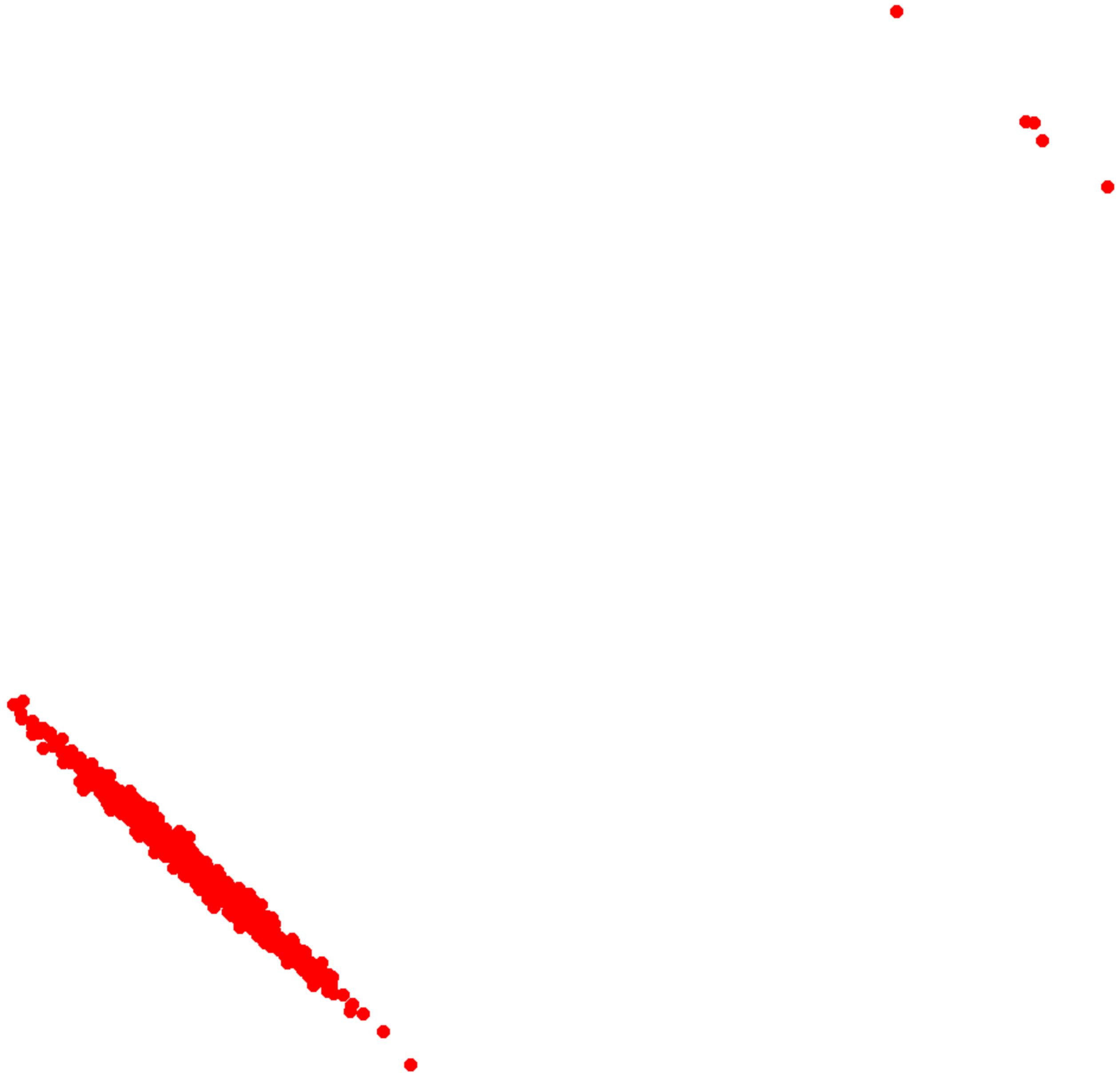
0.05767



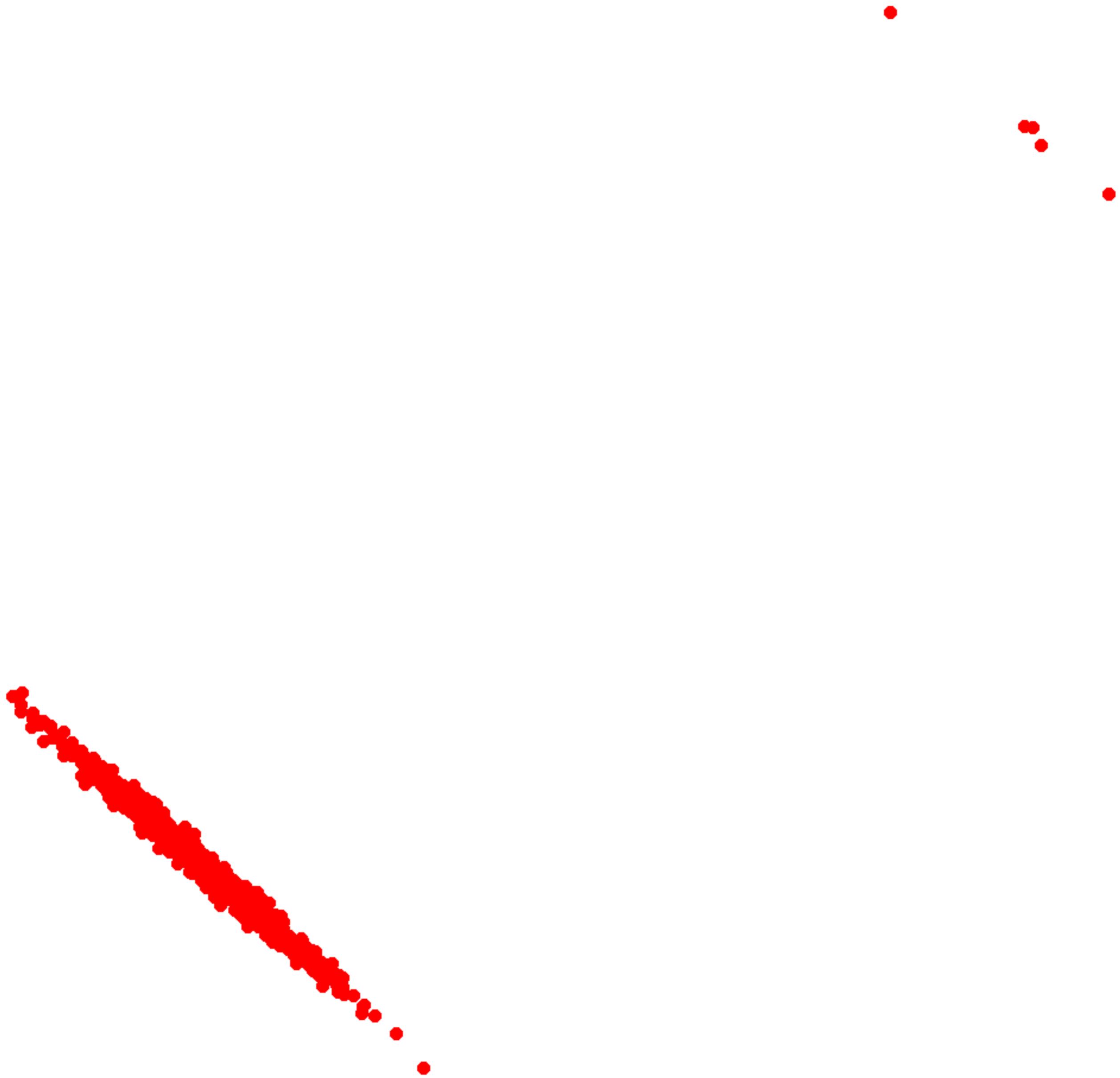
0.01614



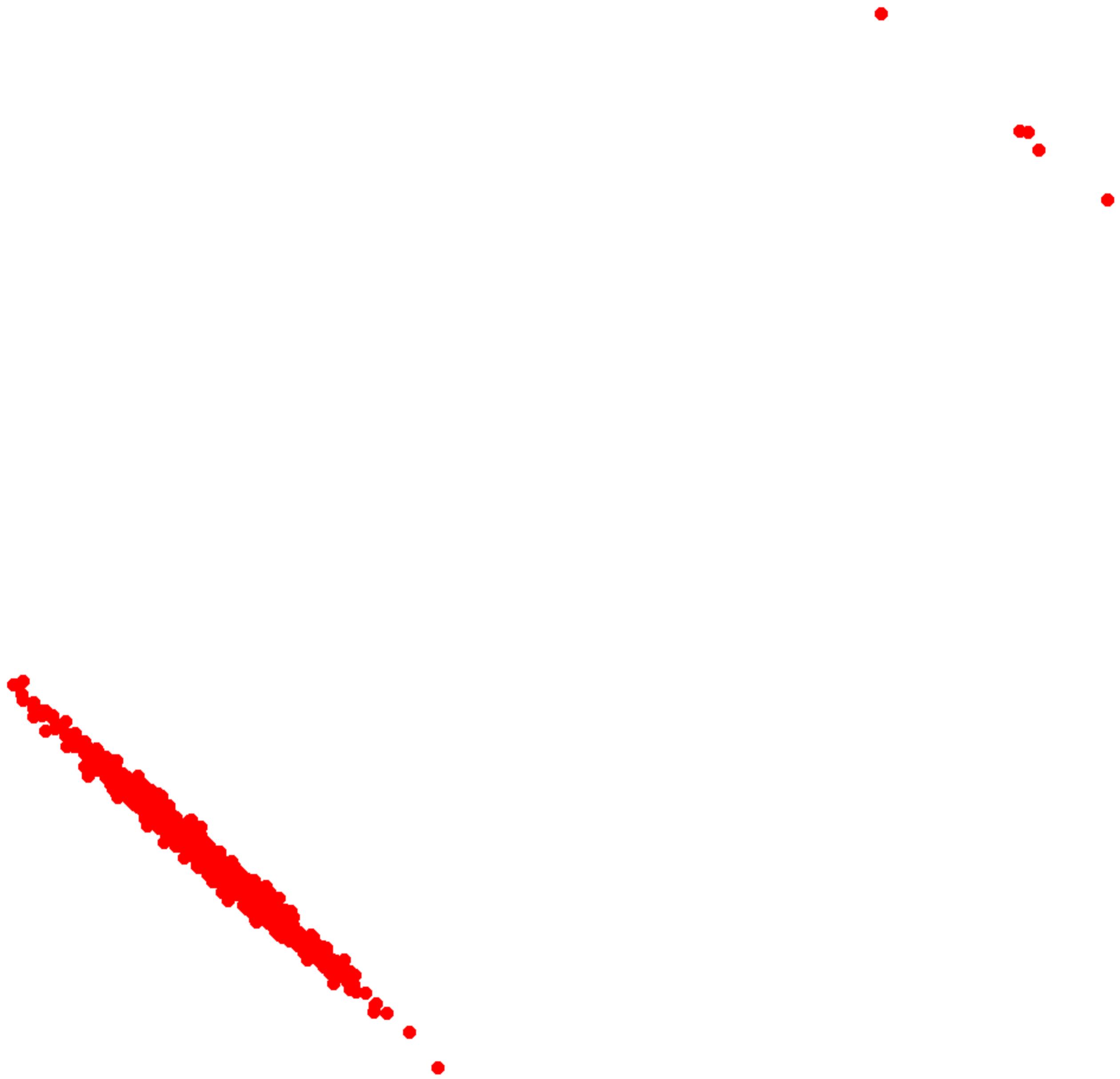
-0.02633



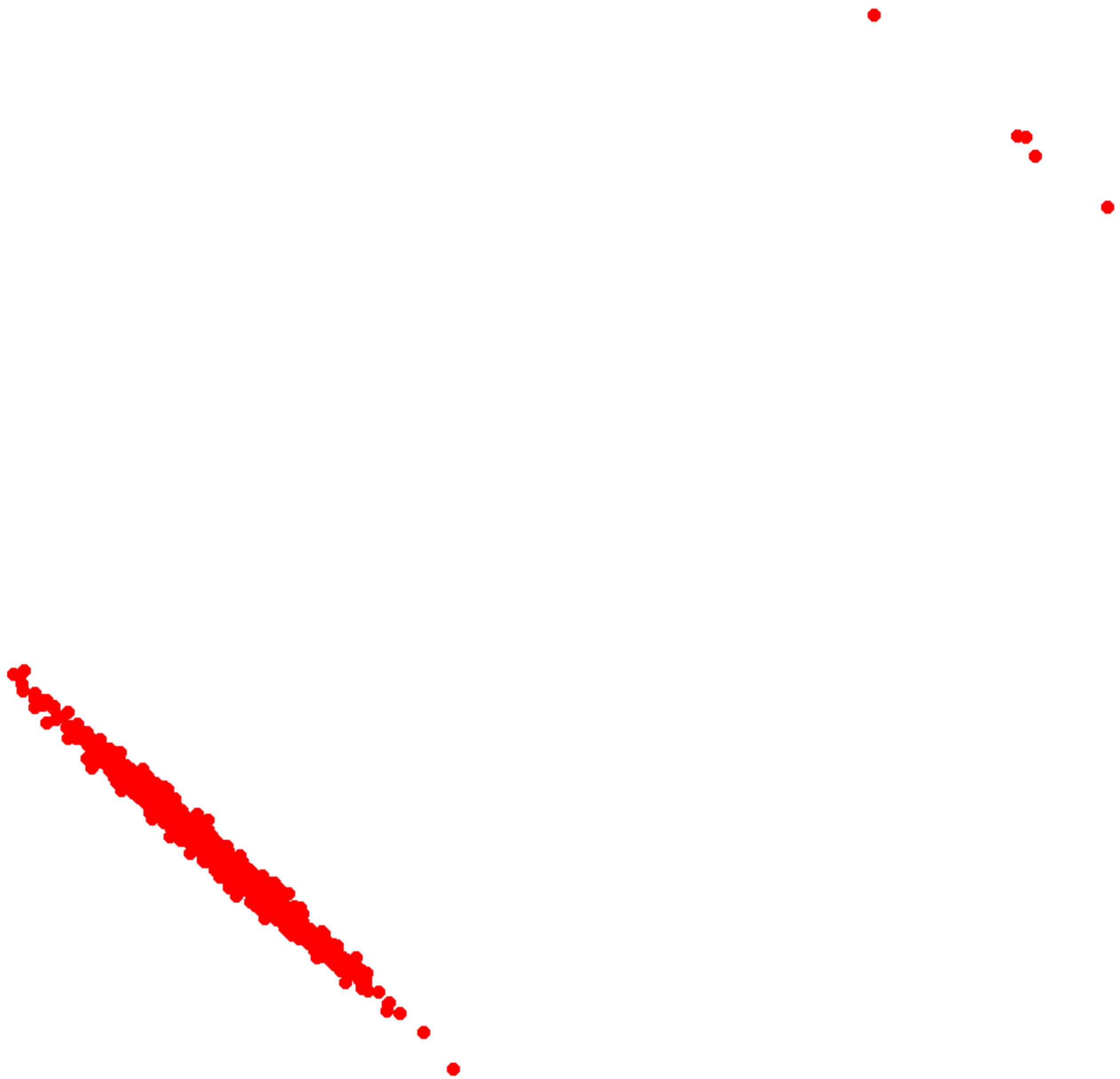
-0.0696



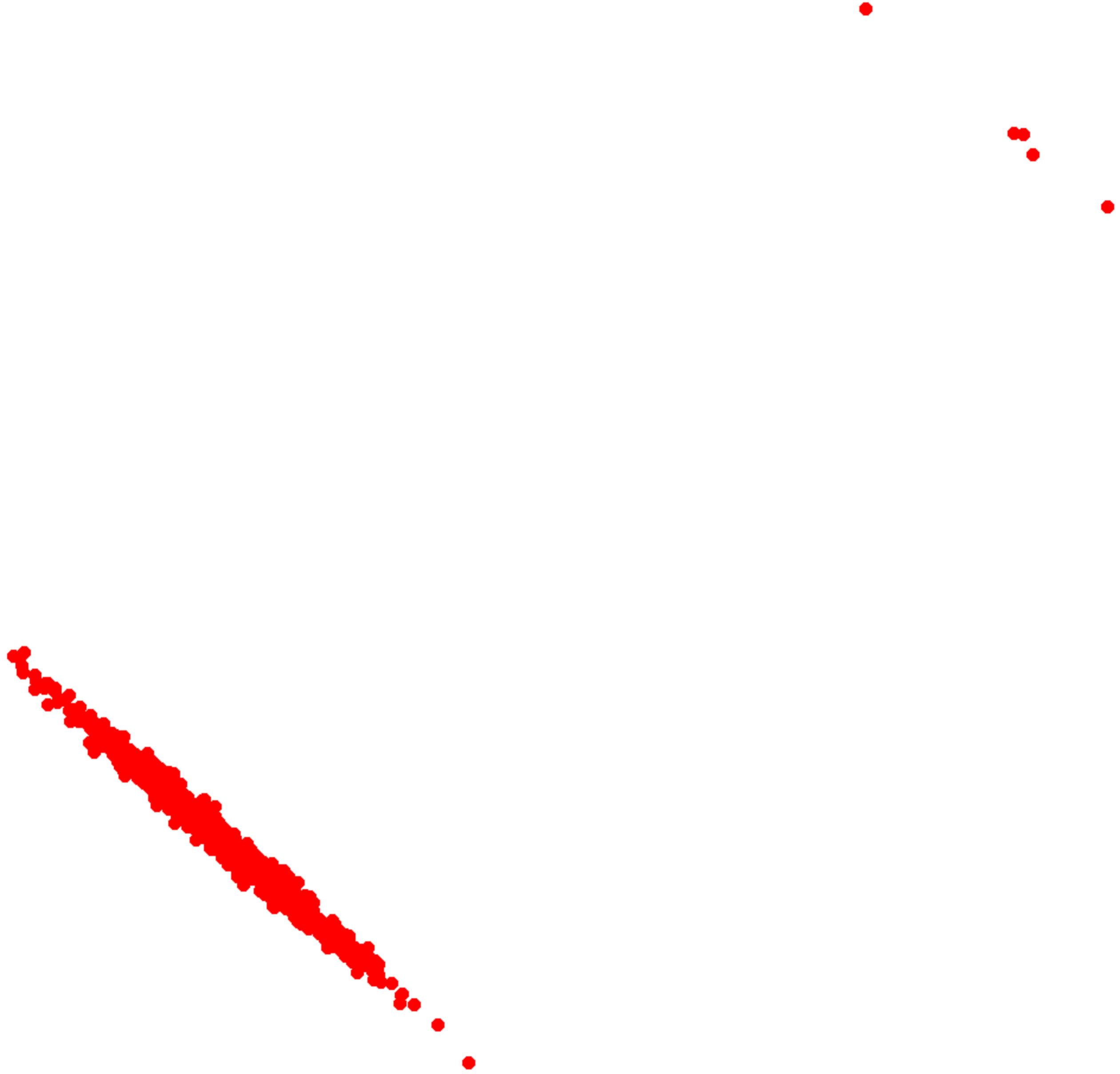
-0.11355



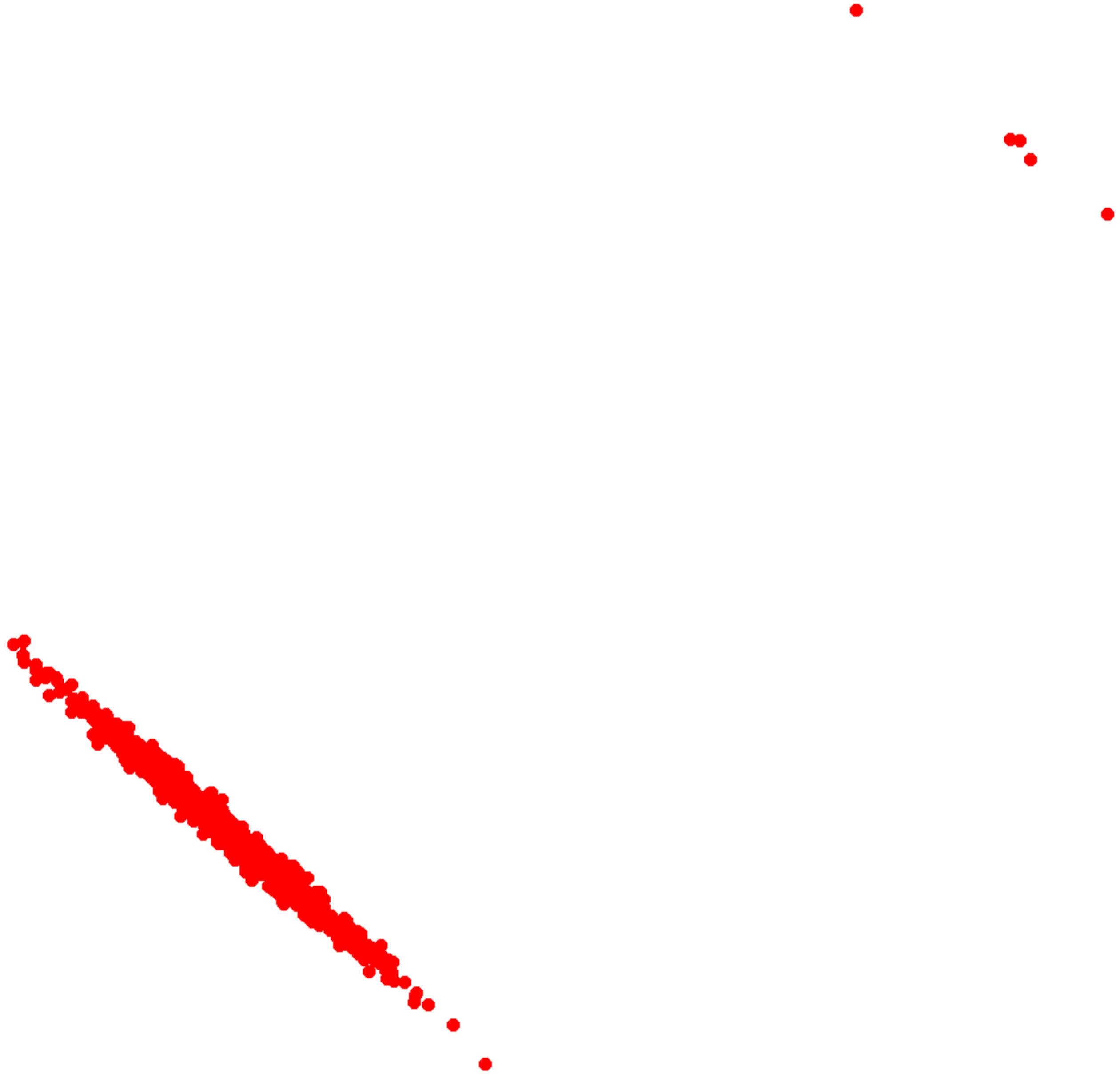
-0.15802



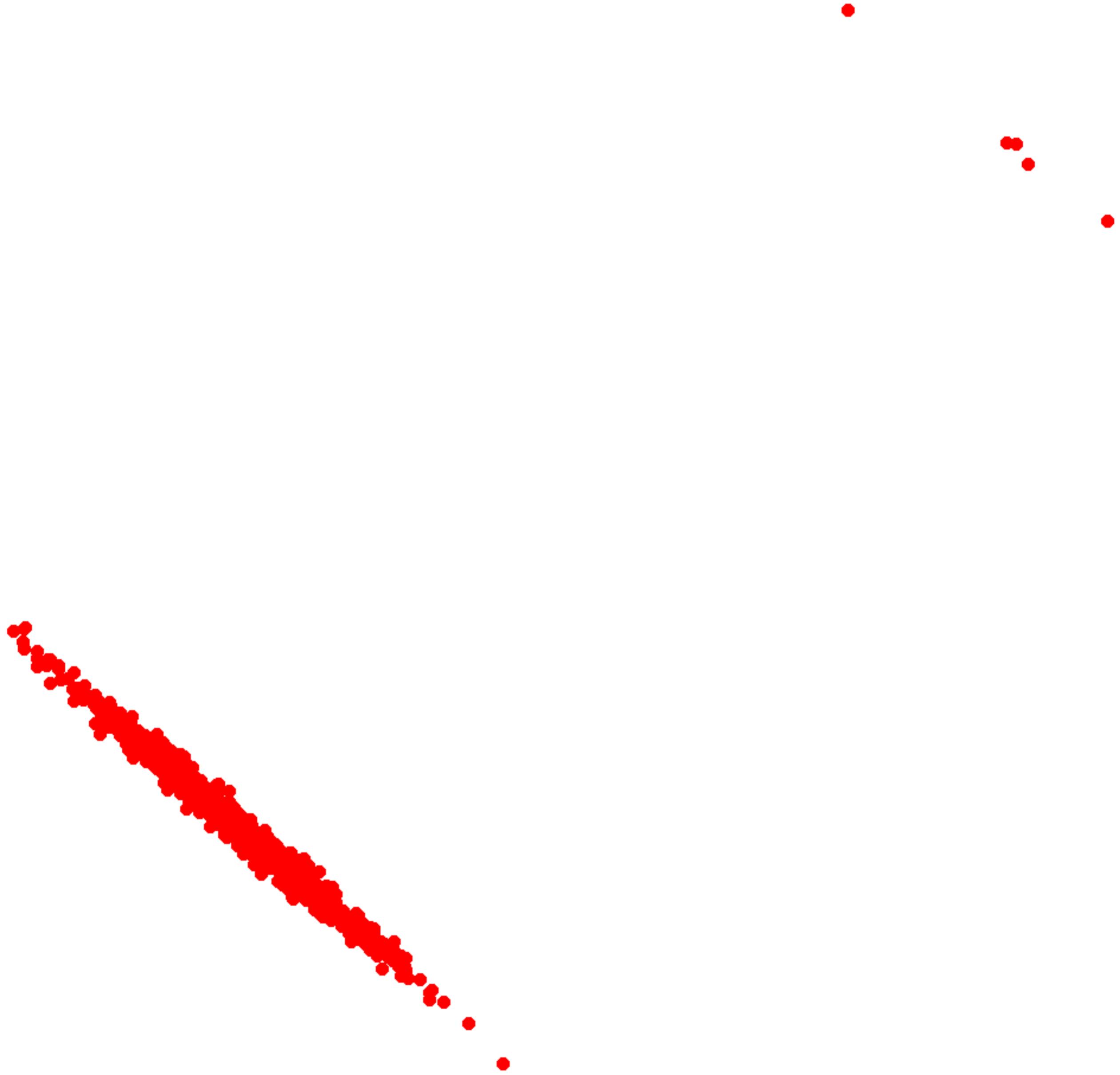
-0.20285



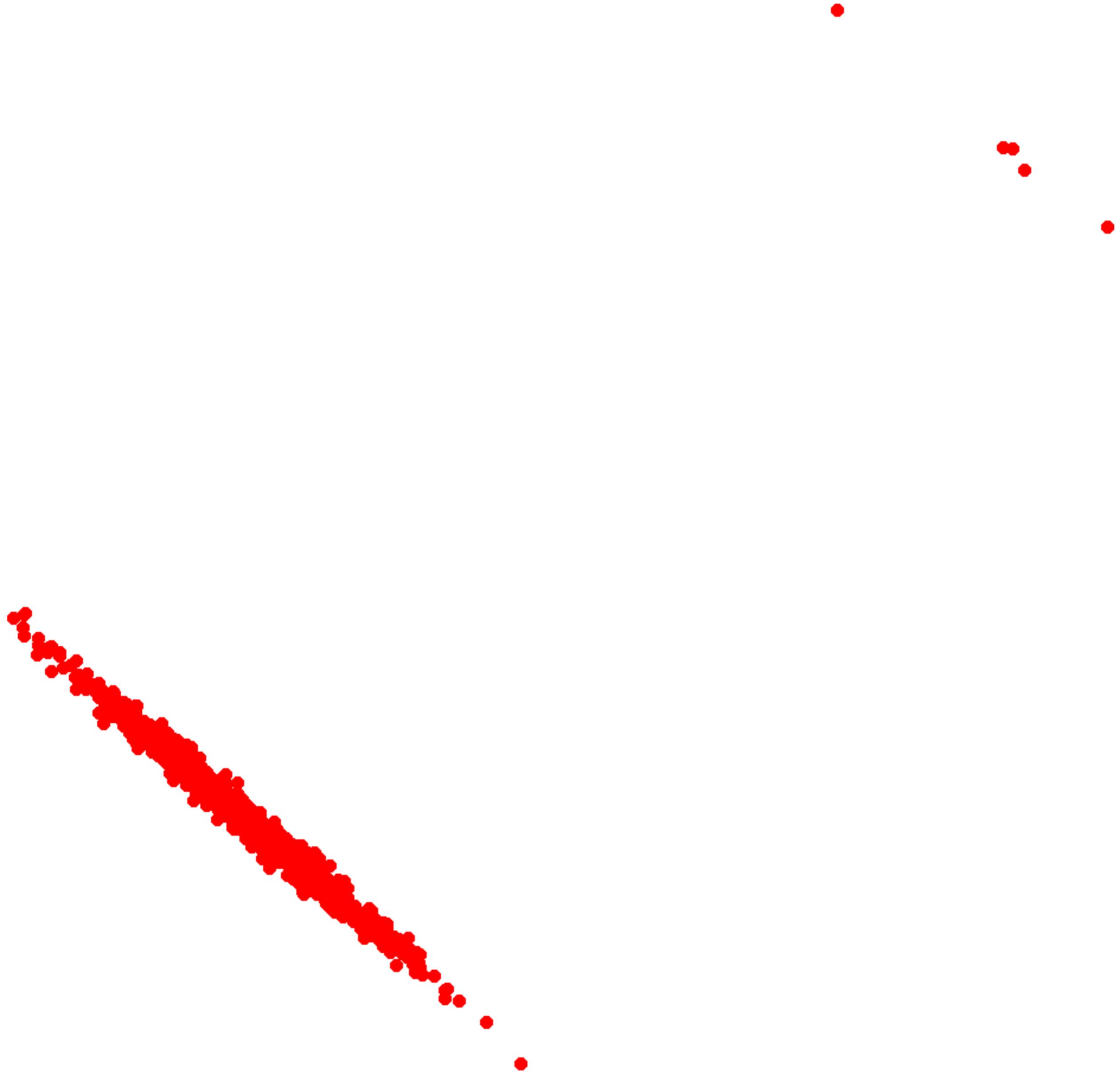
-0.24786



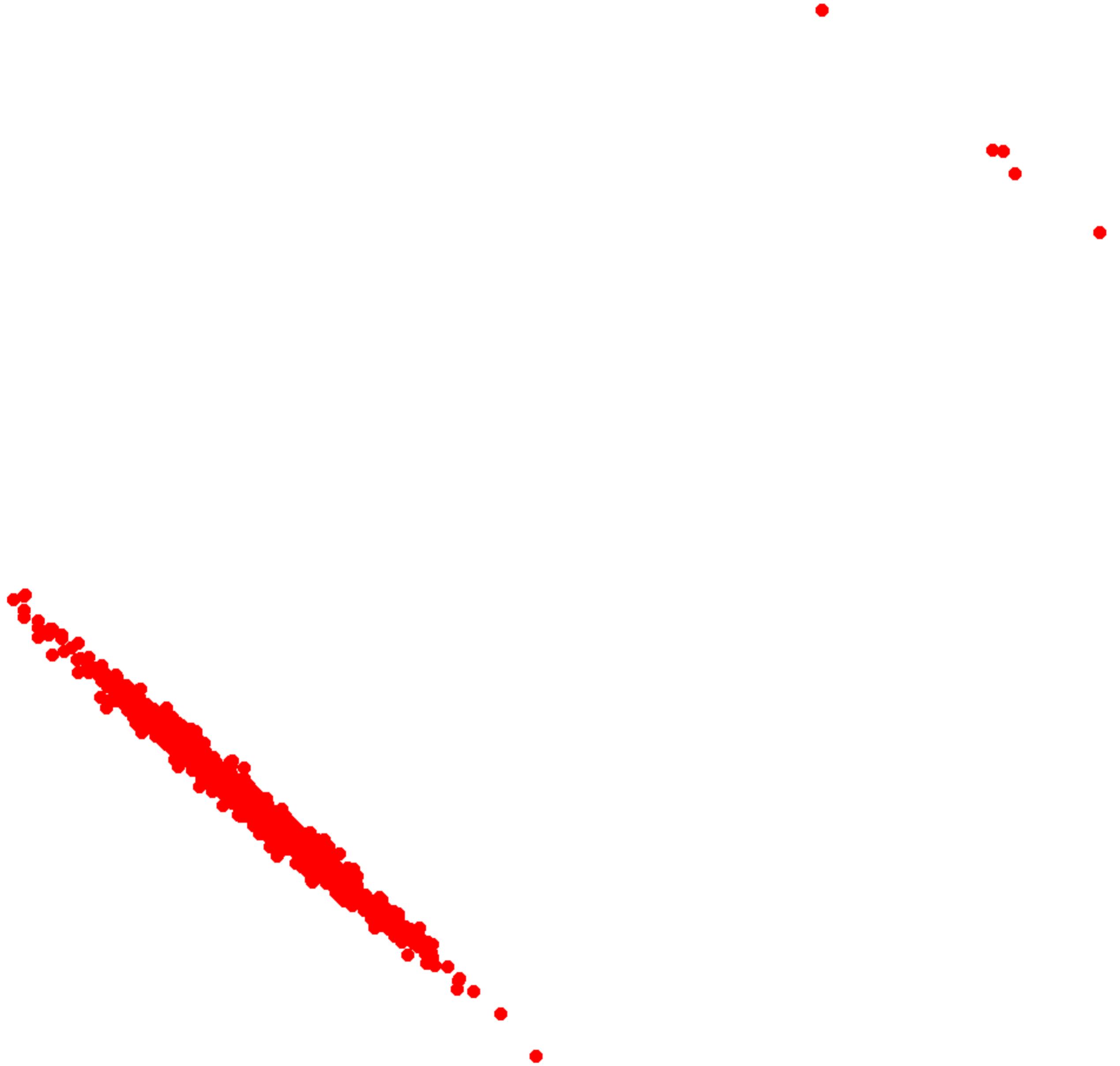
-0.29286



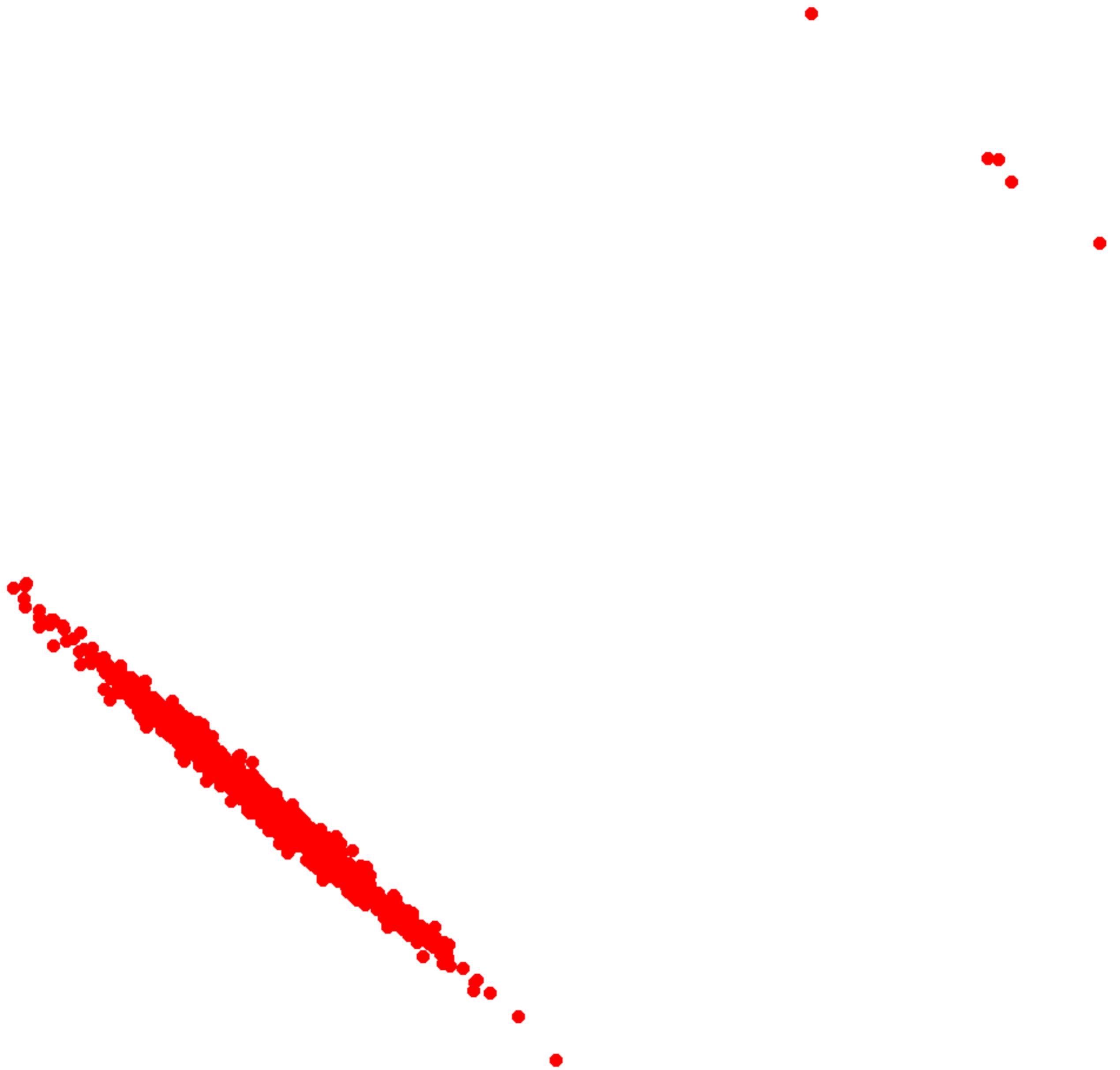
-0.33763



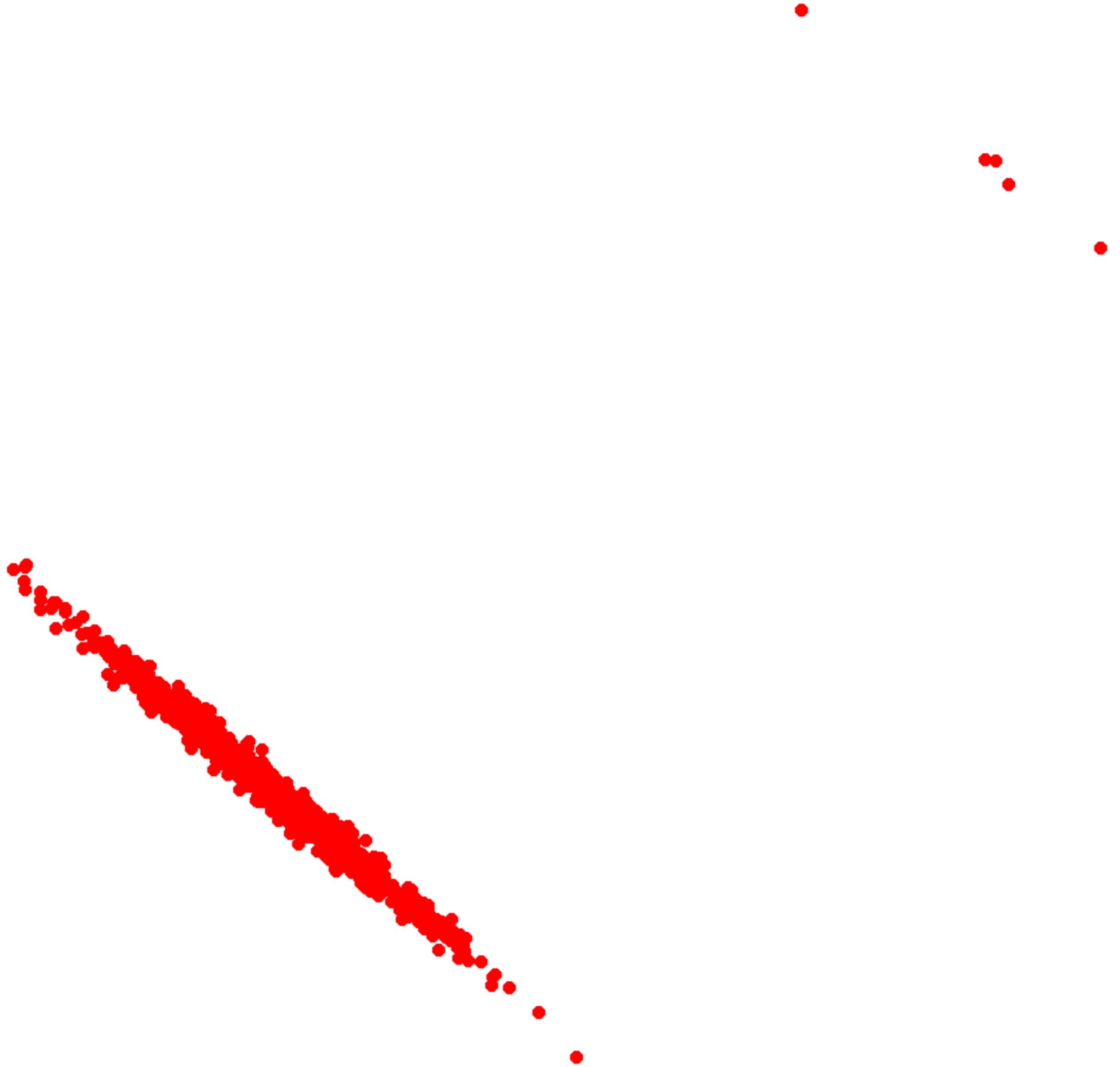
-0.38199



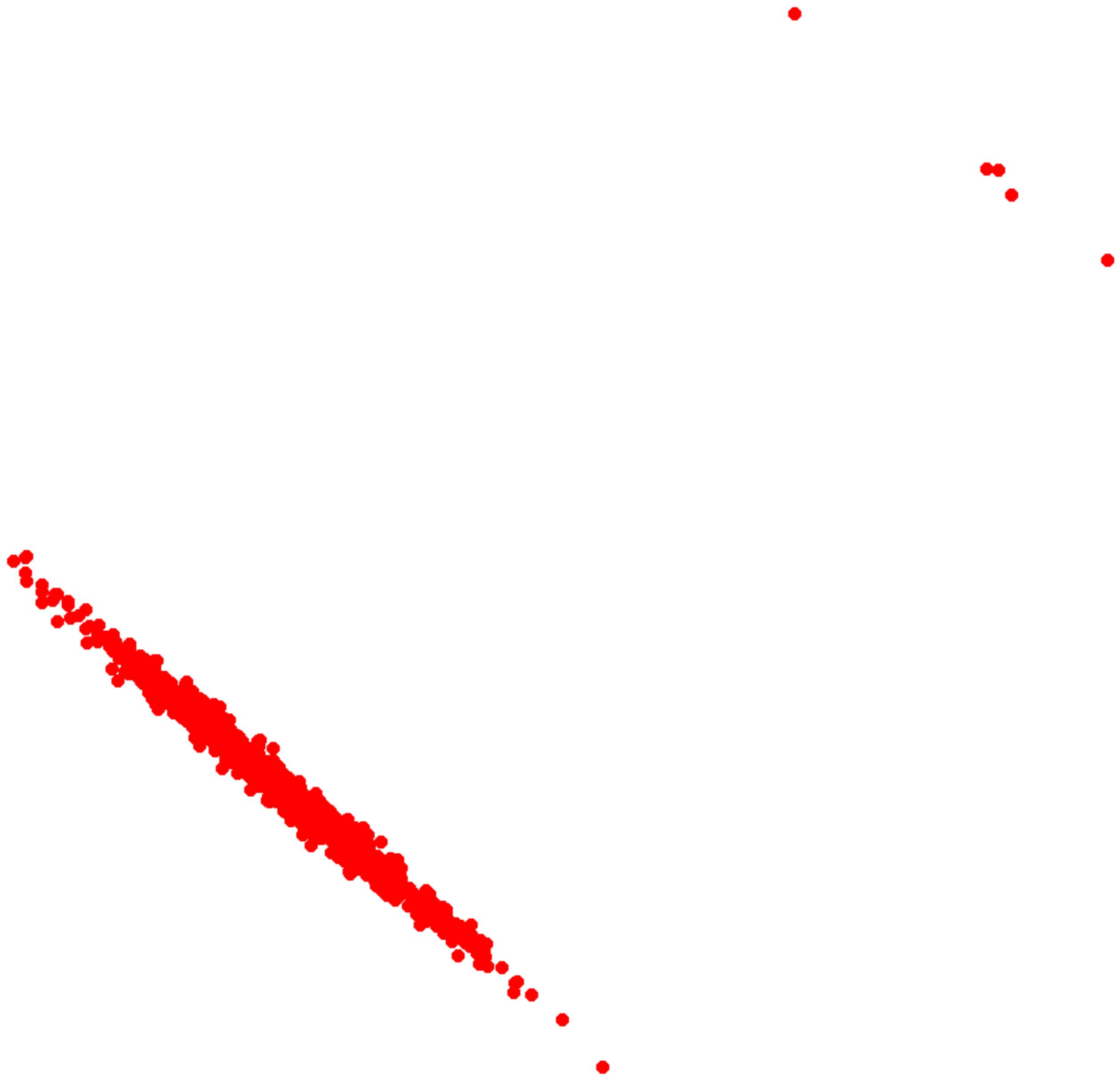
-0.42571



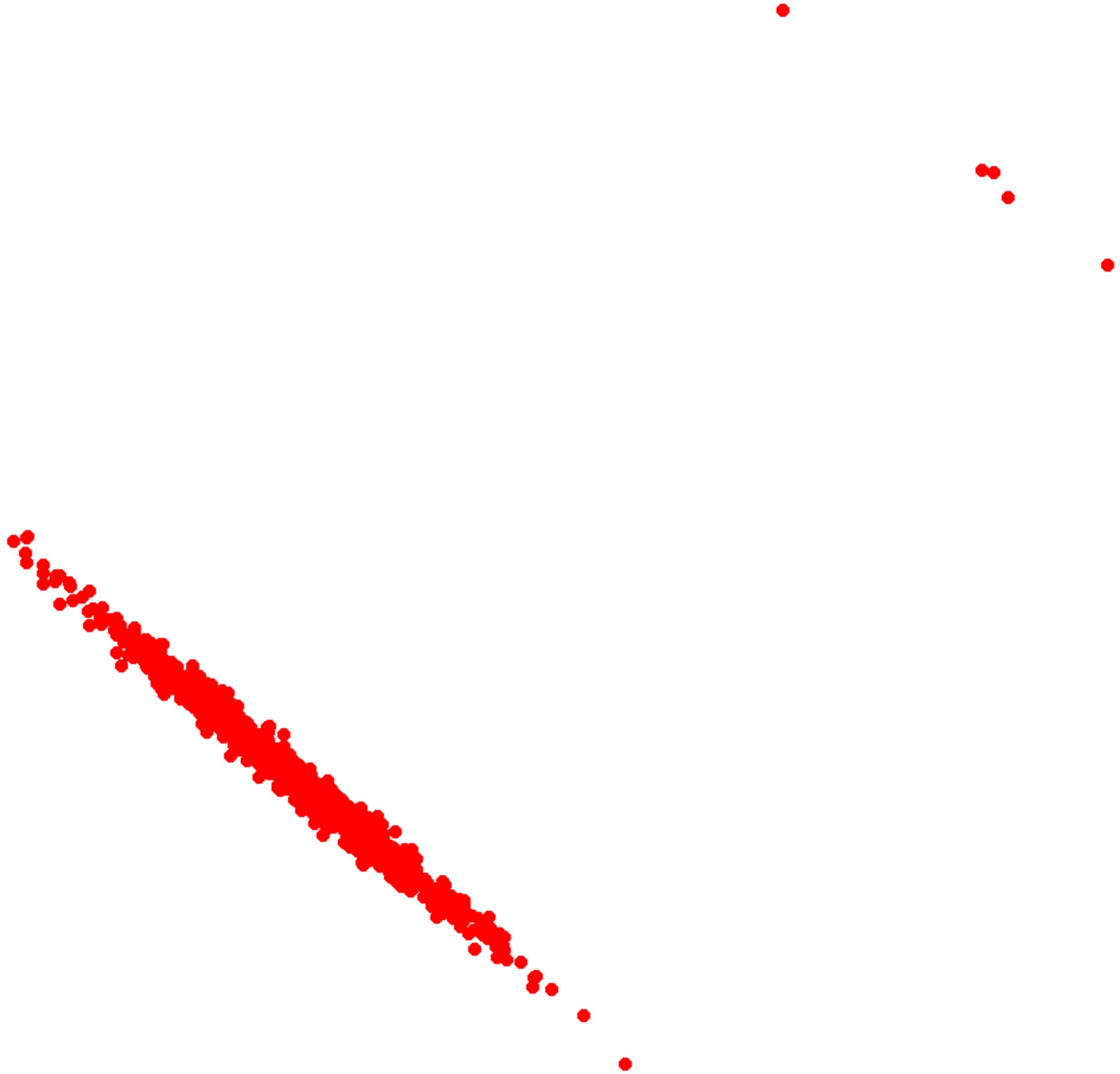
-0.46858



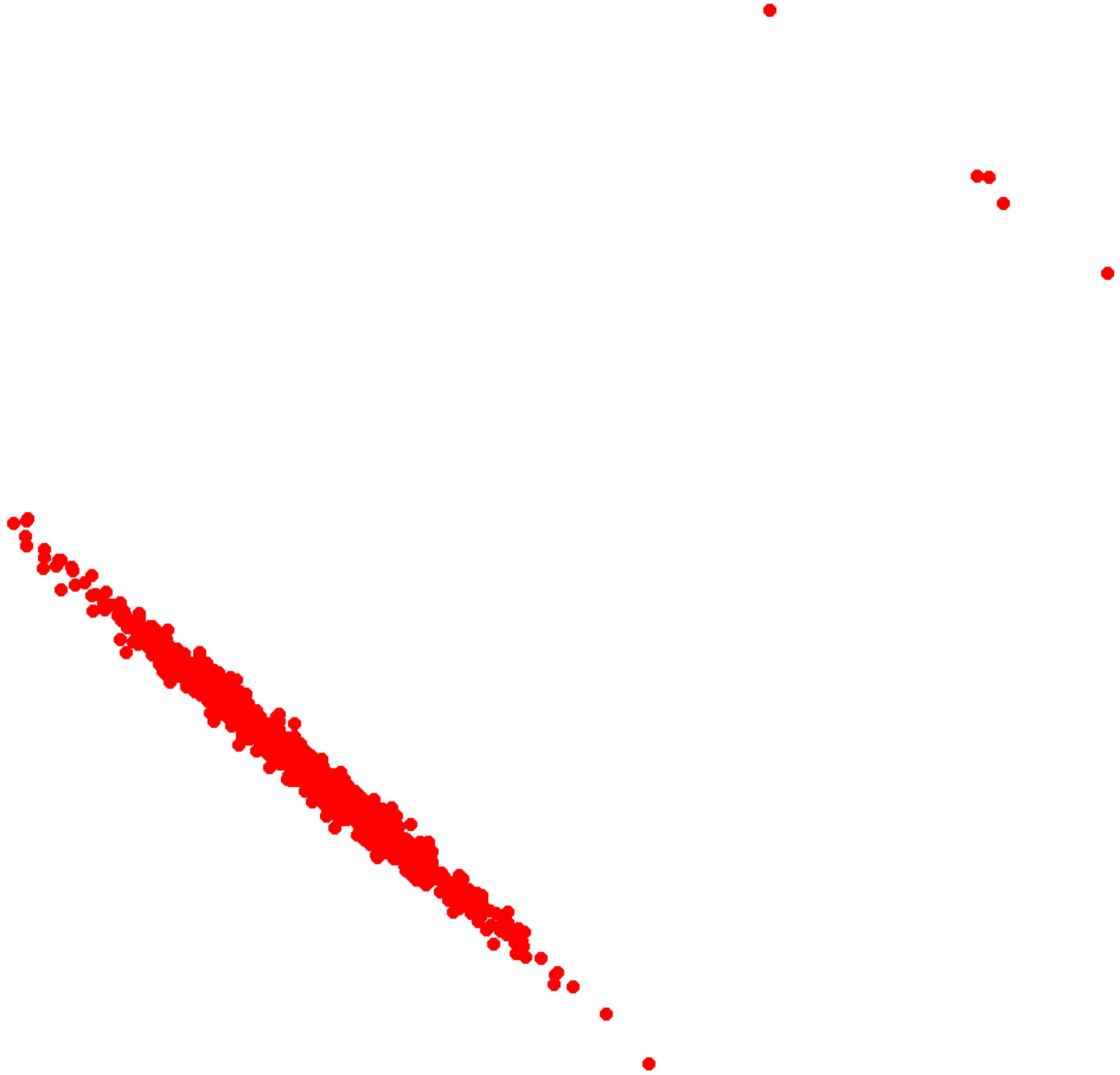
-0.51041



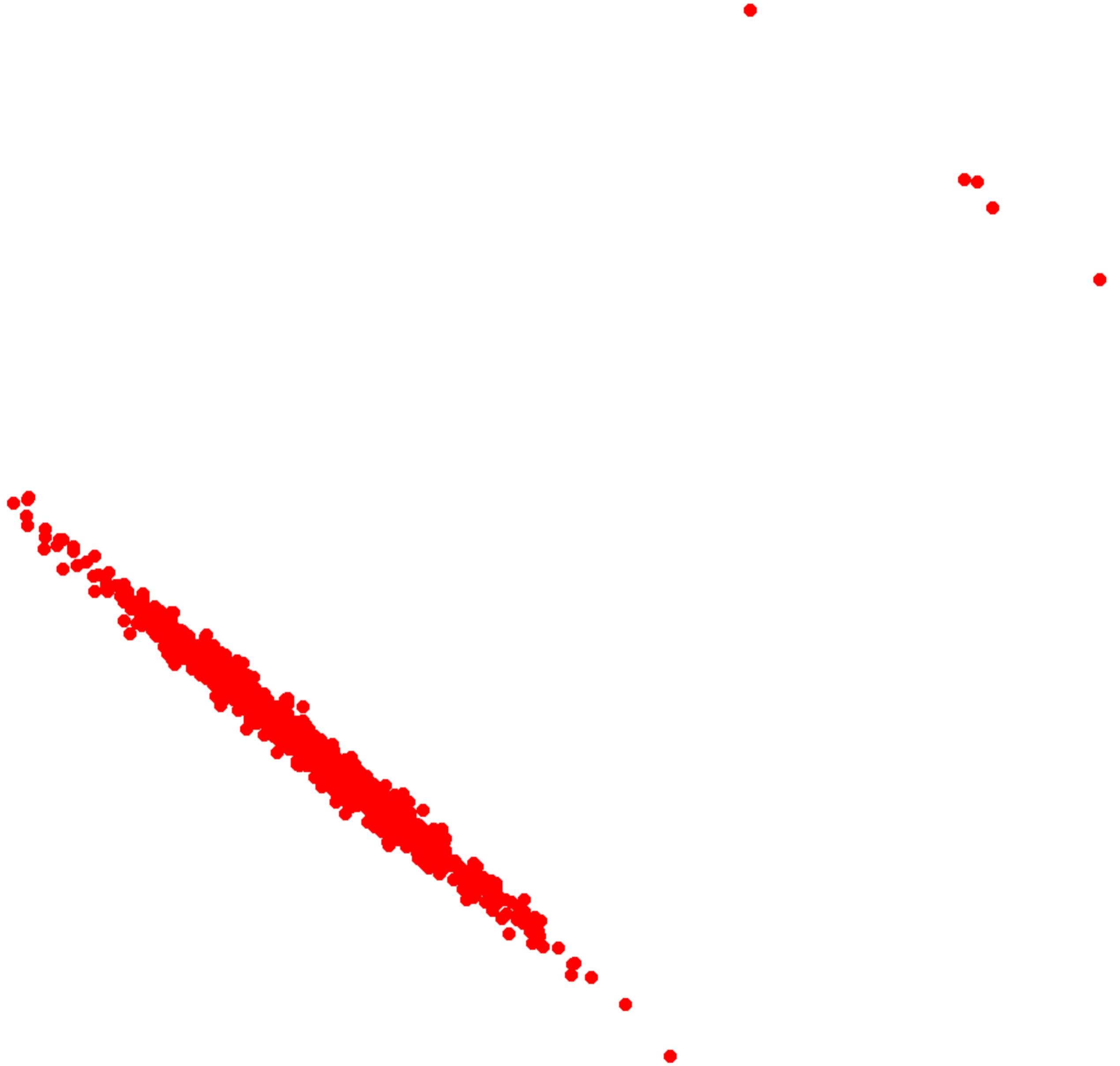
-0.55099



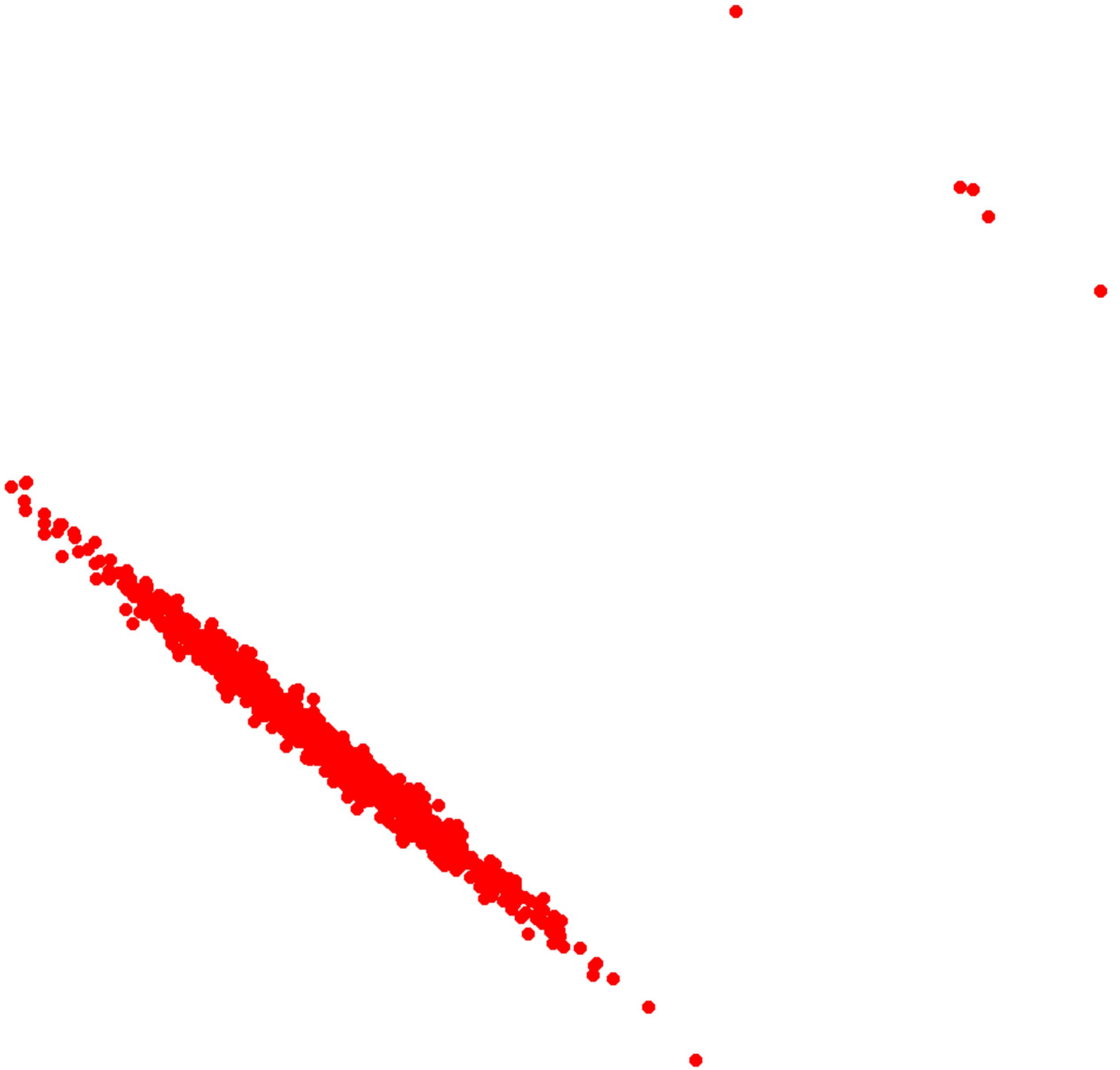
-0.59014



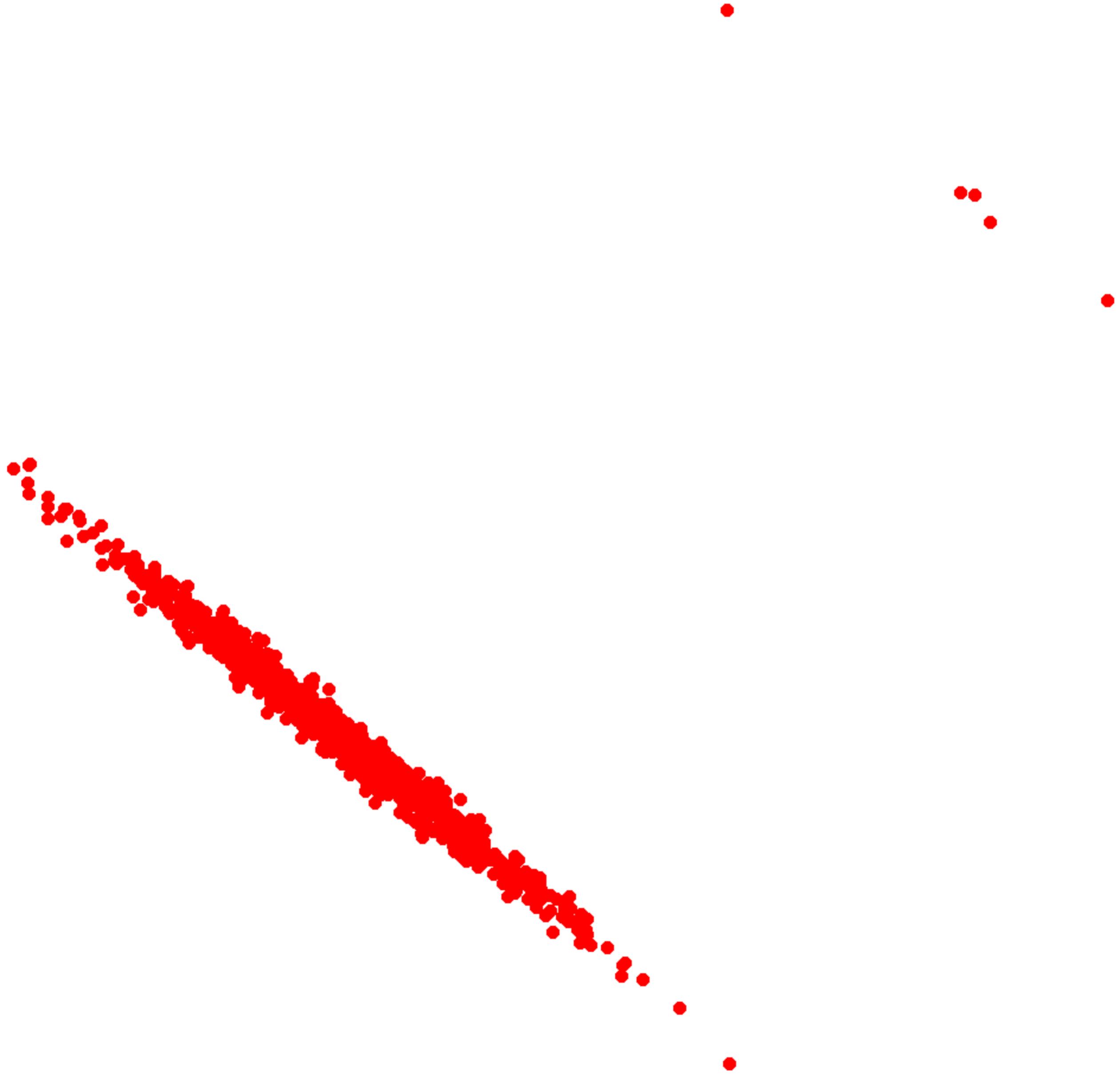
-0.62769



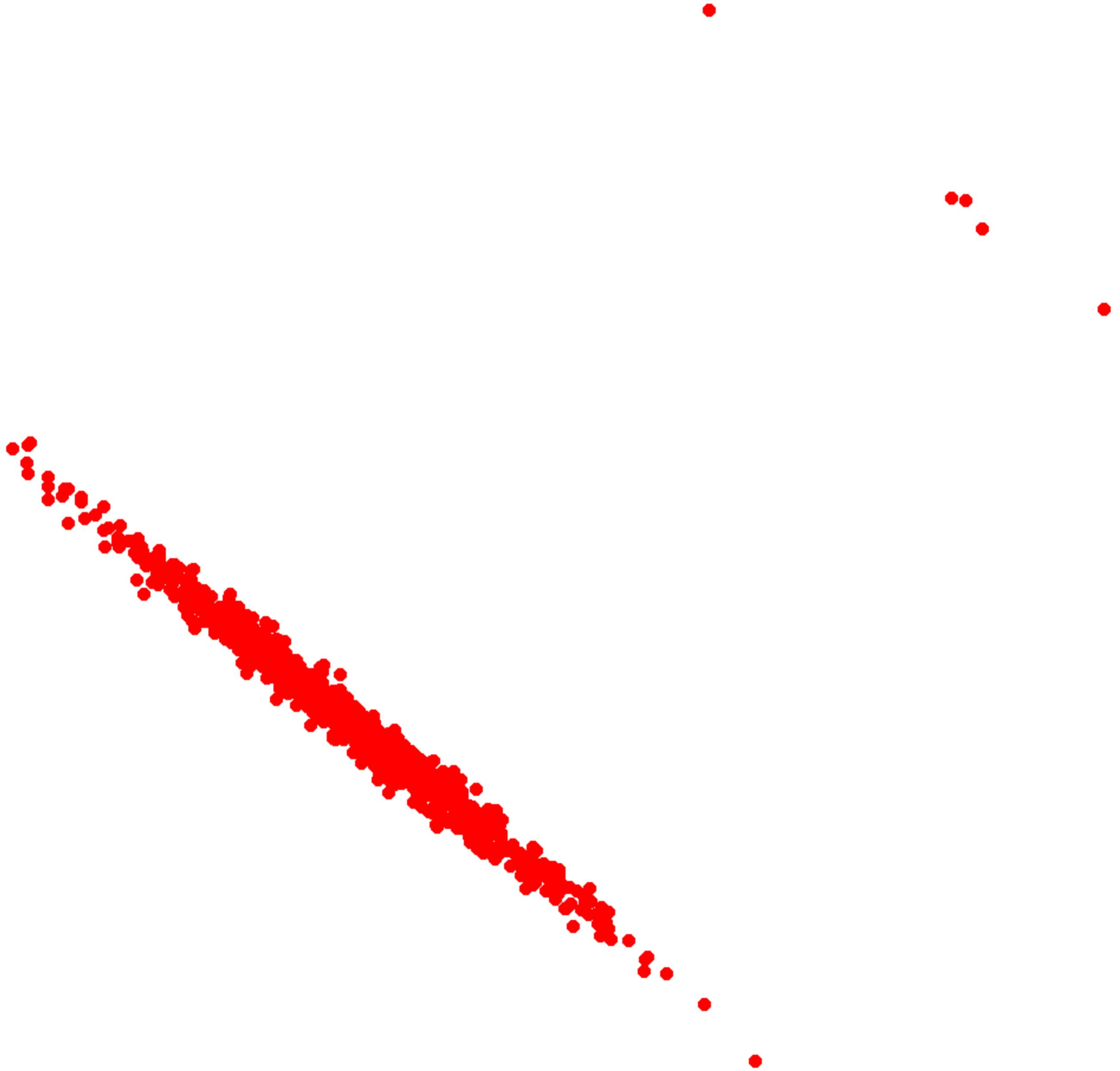
-0.6635



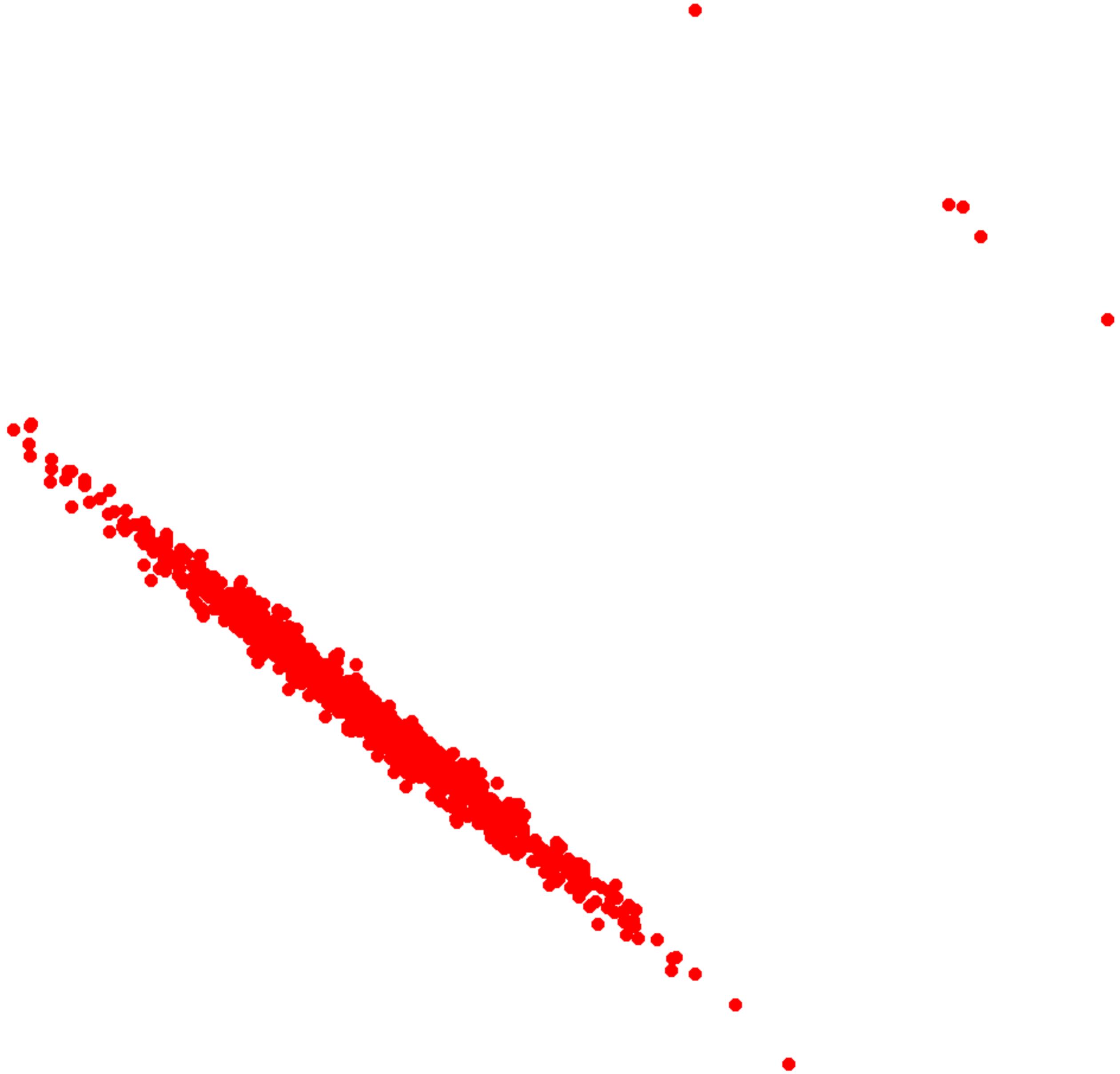
-0.69743



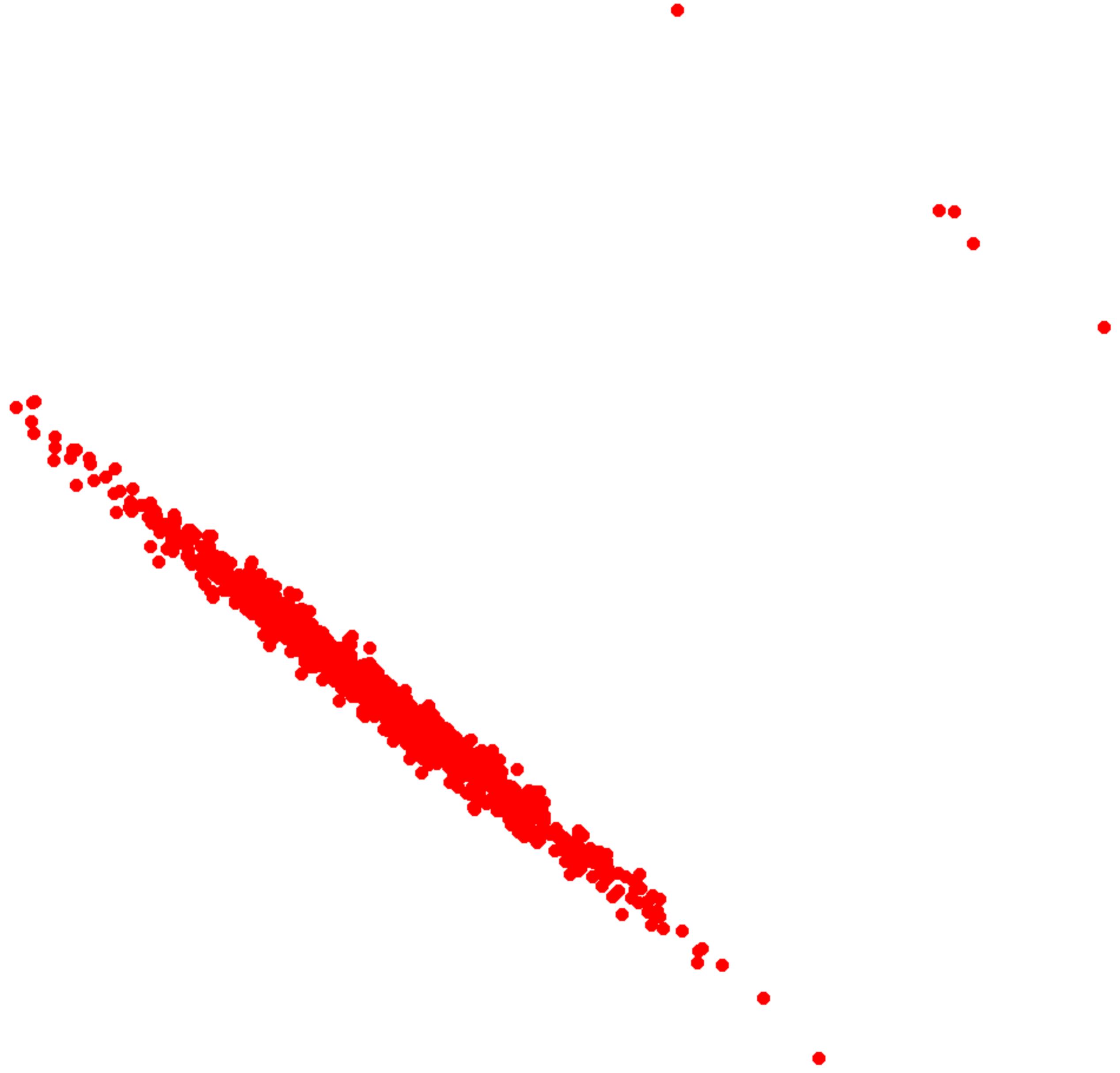
-0.72939



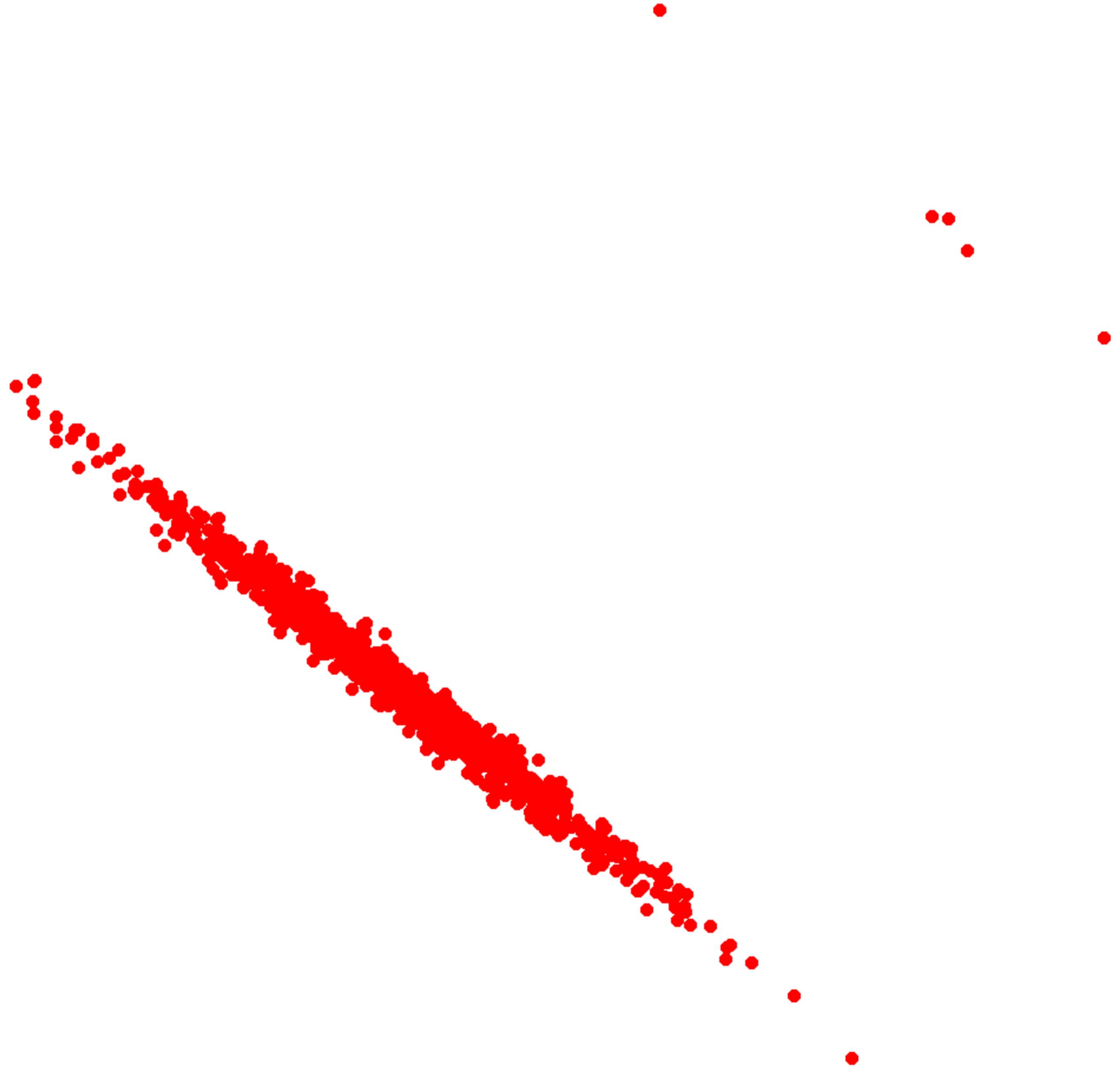
-0.75929



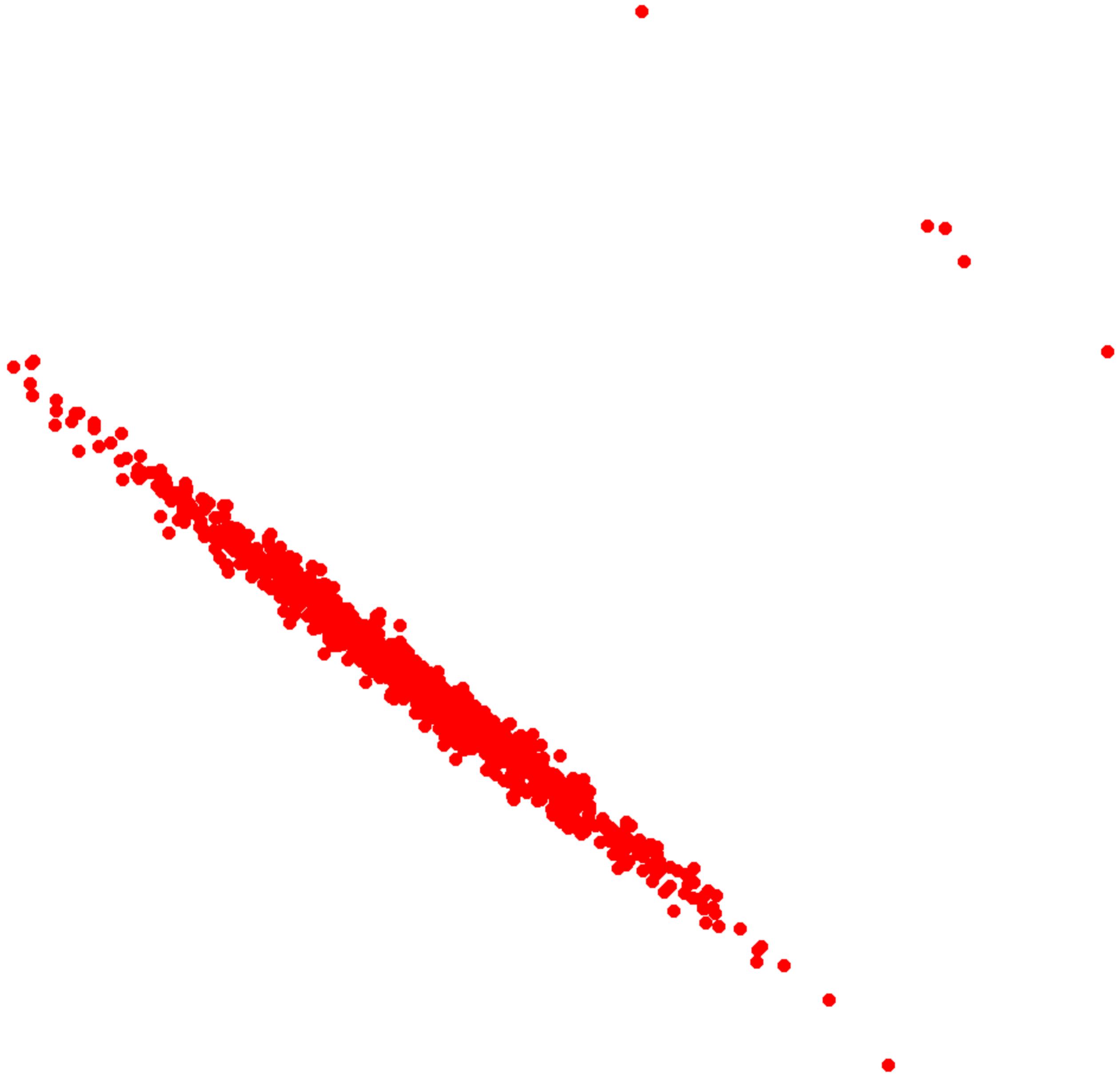
-0.78709



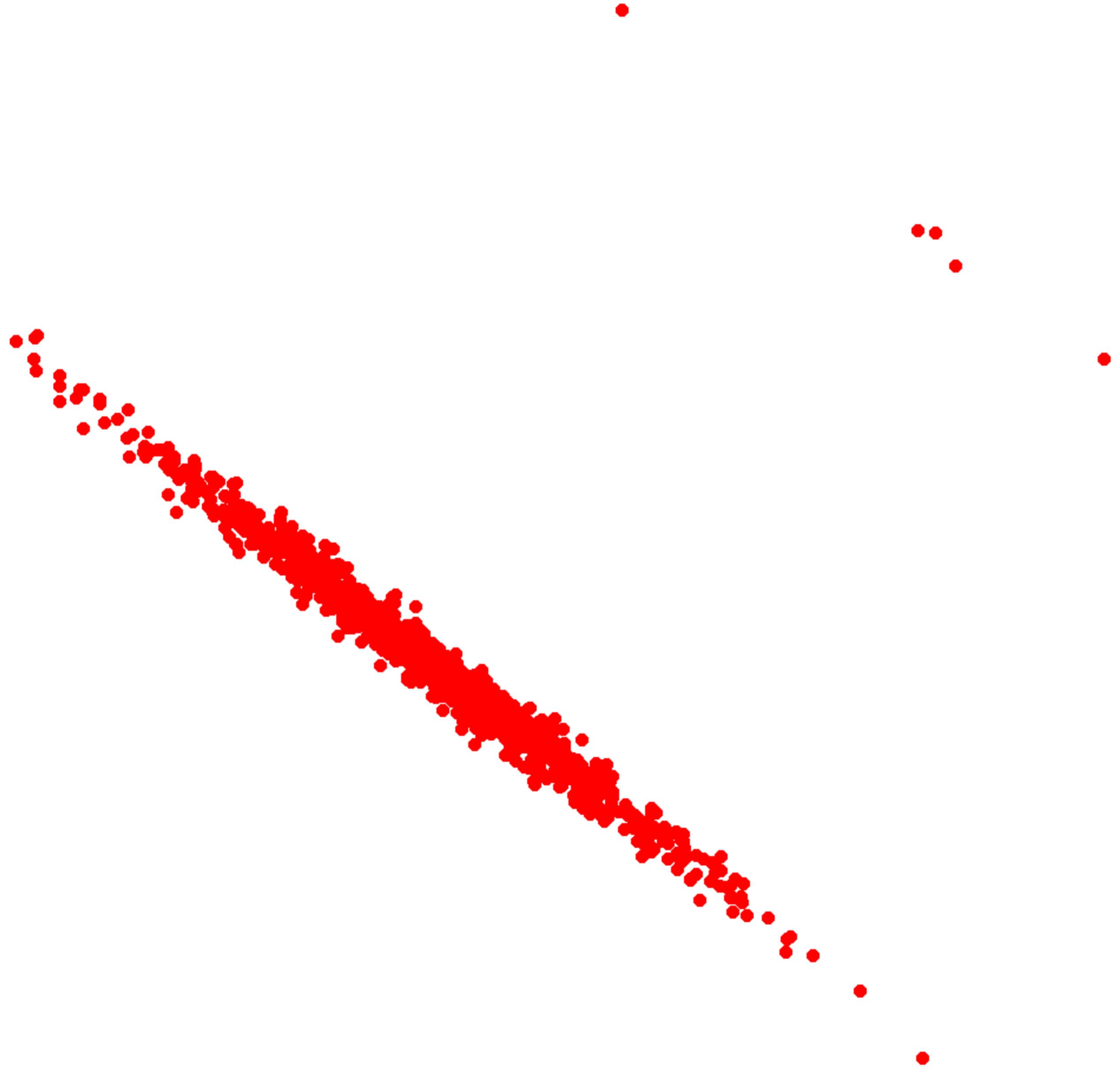
-0.81275



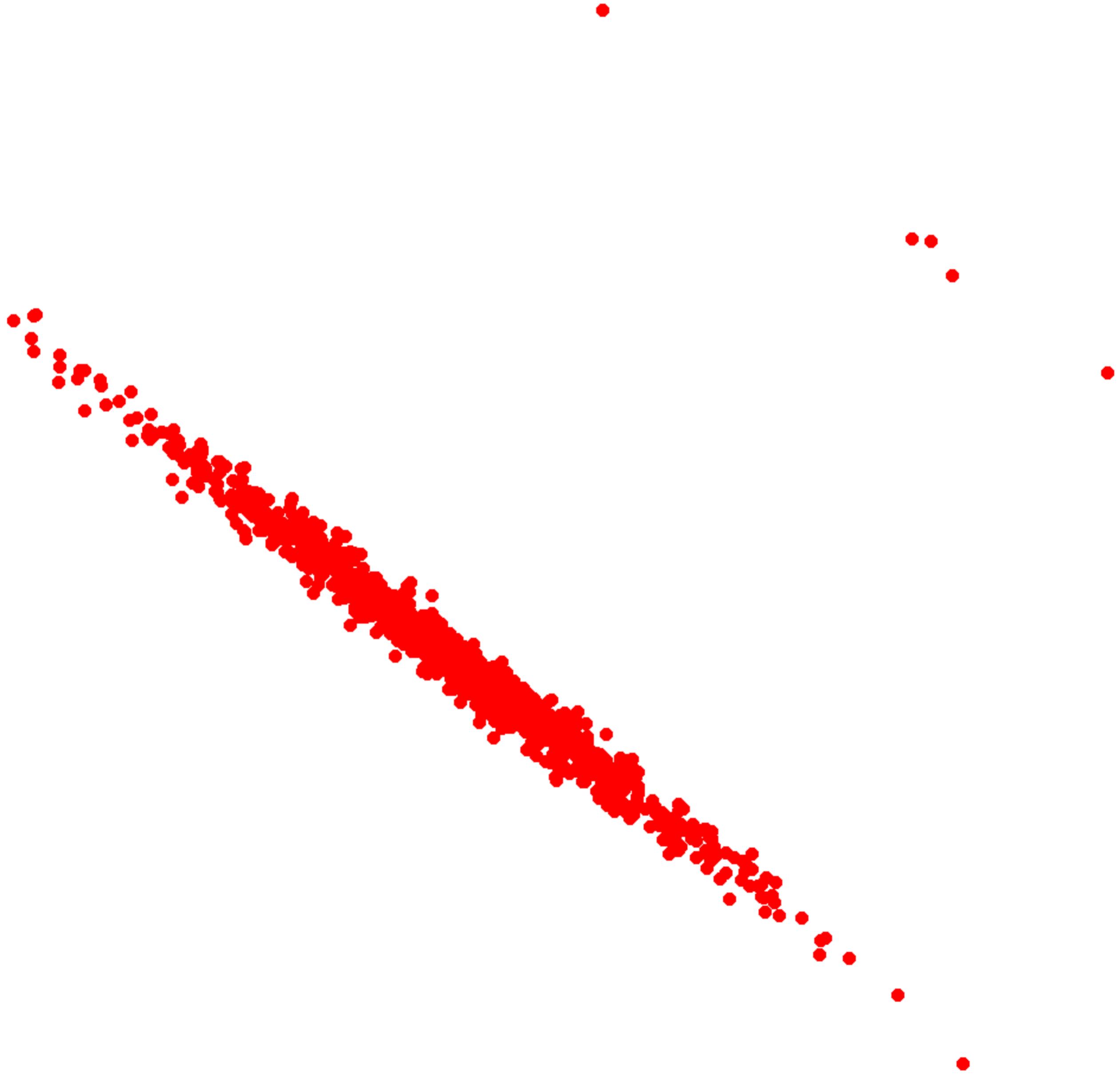
-0.83628



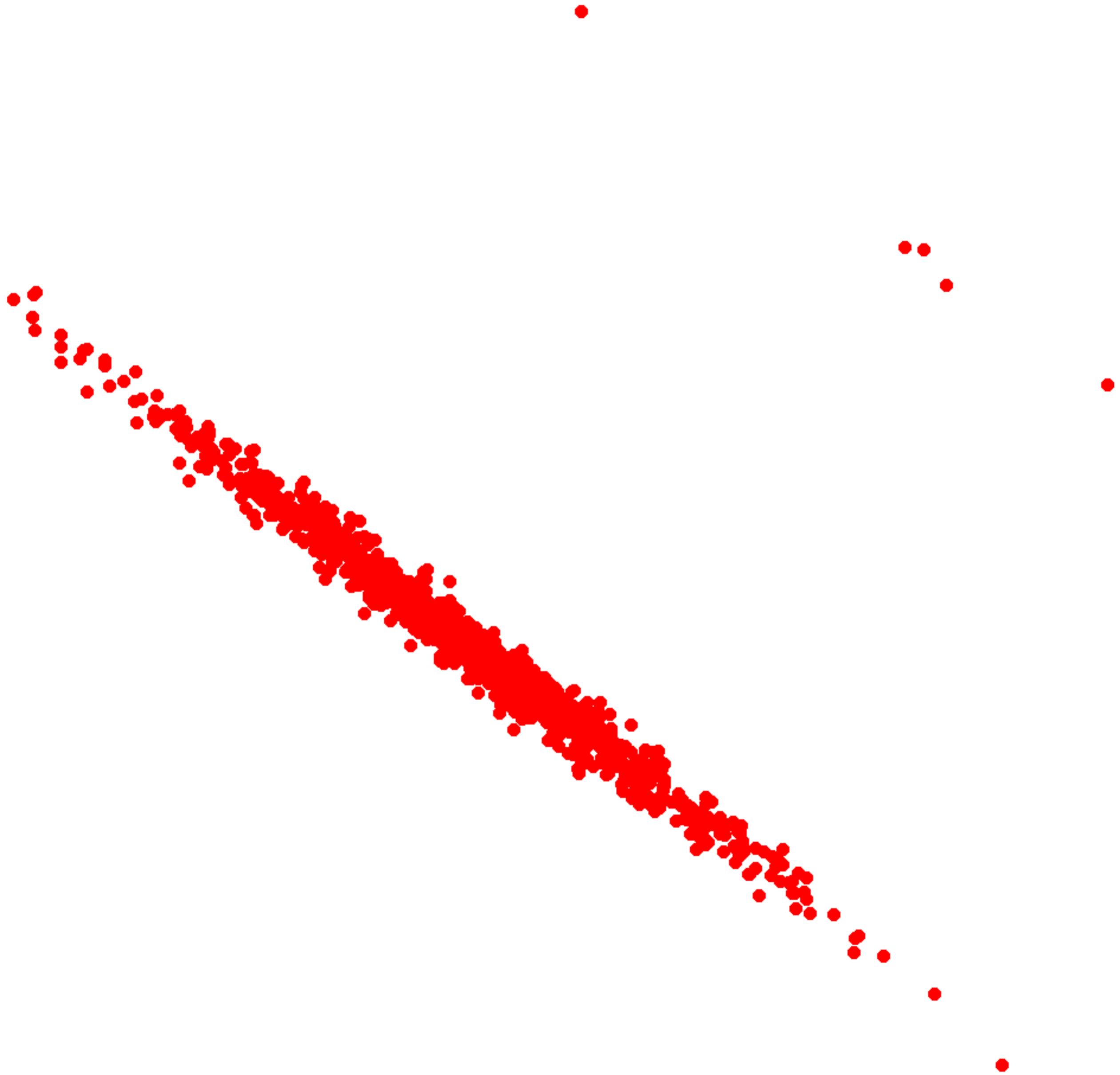
-0.8577



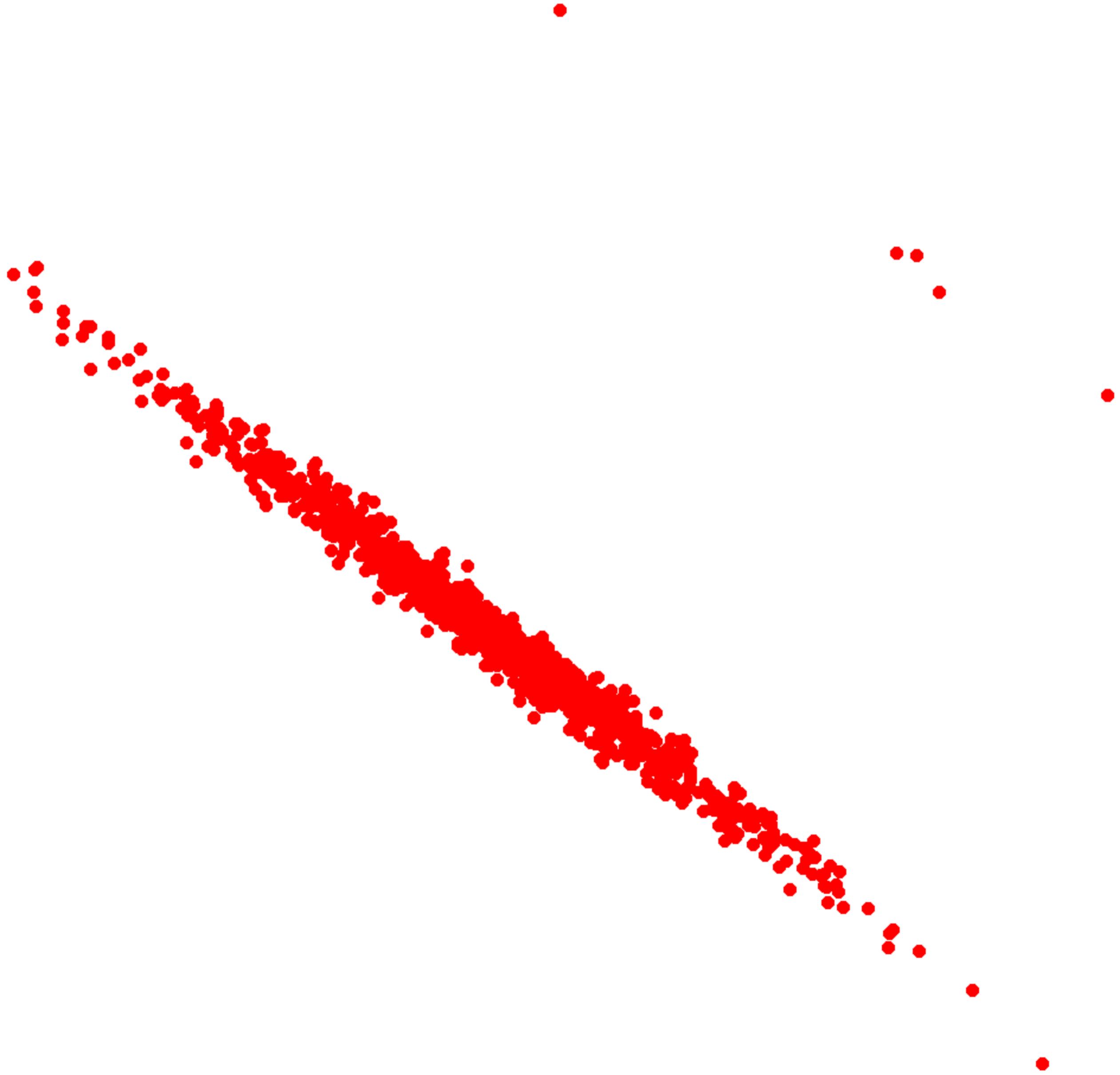
-0.87706



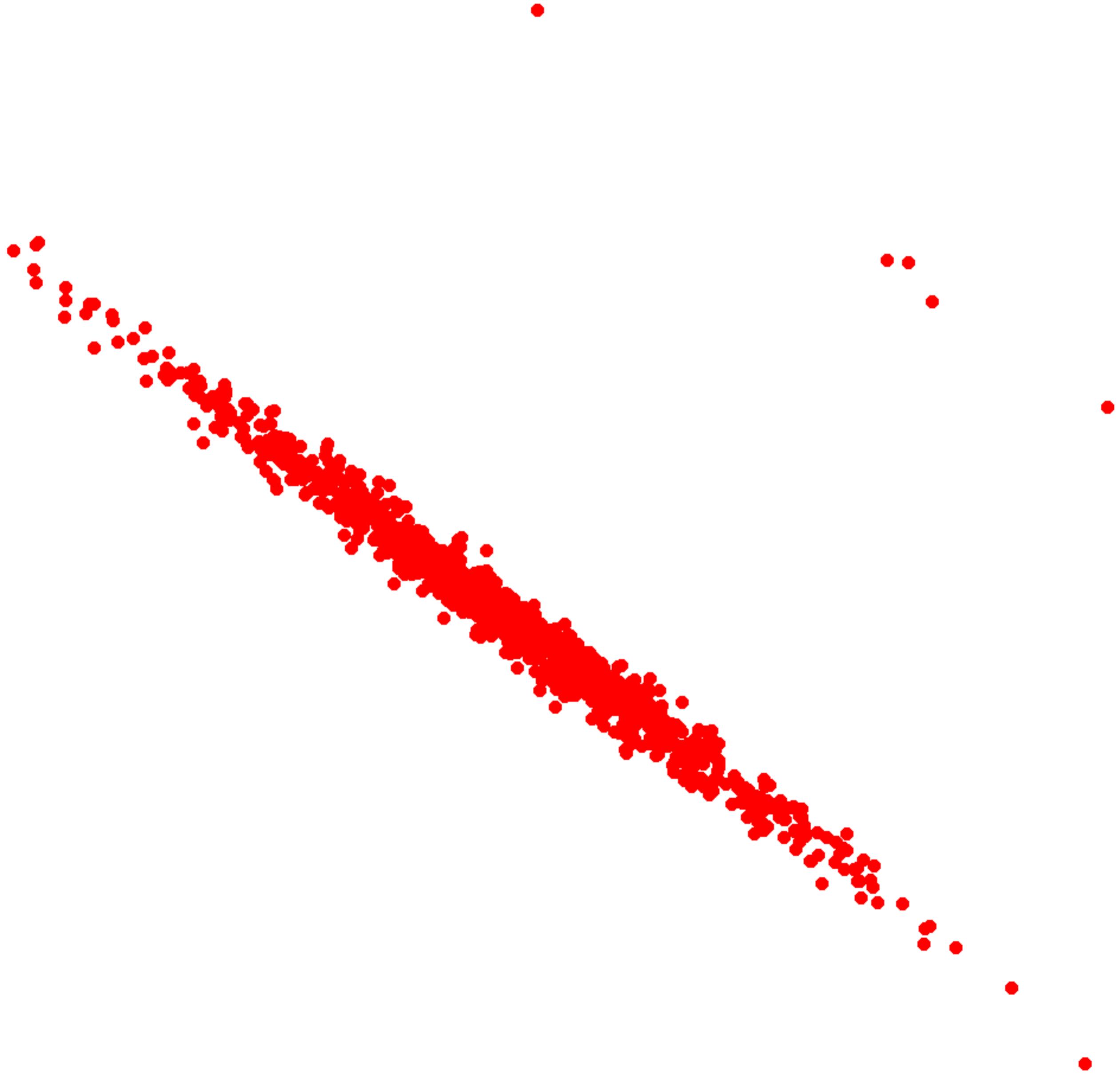
-0.89441



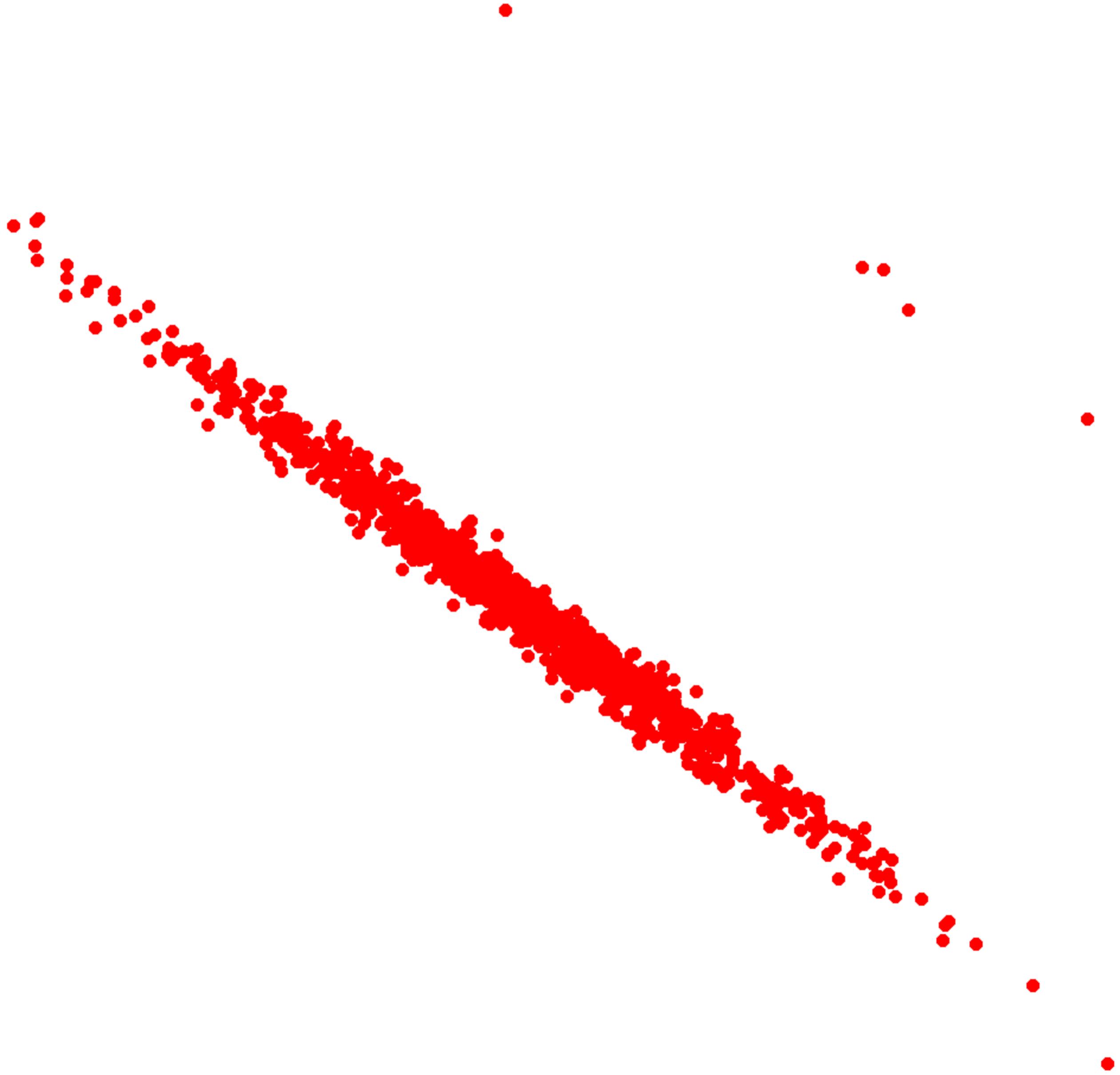
-0.90985



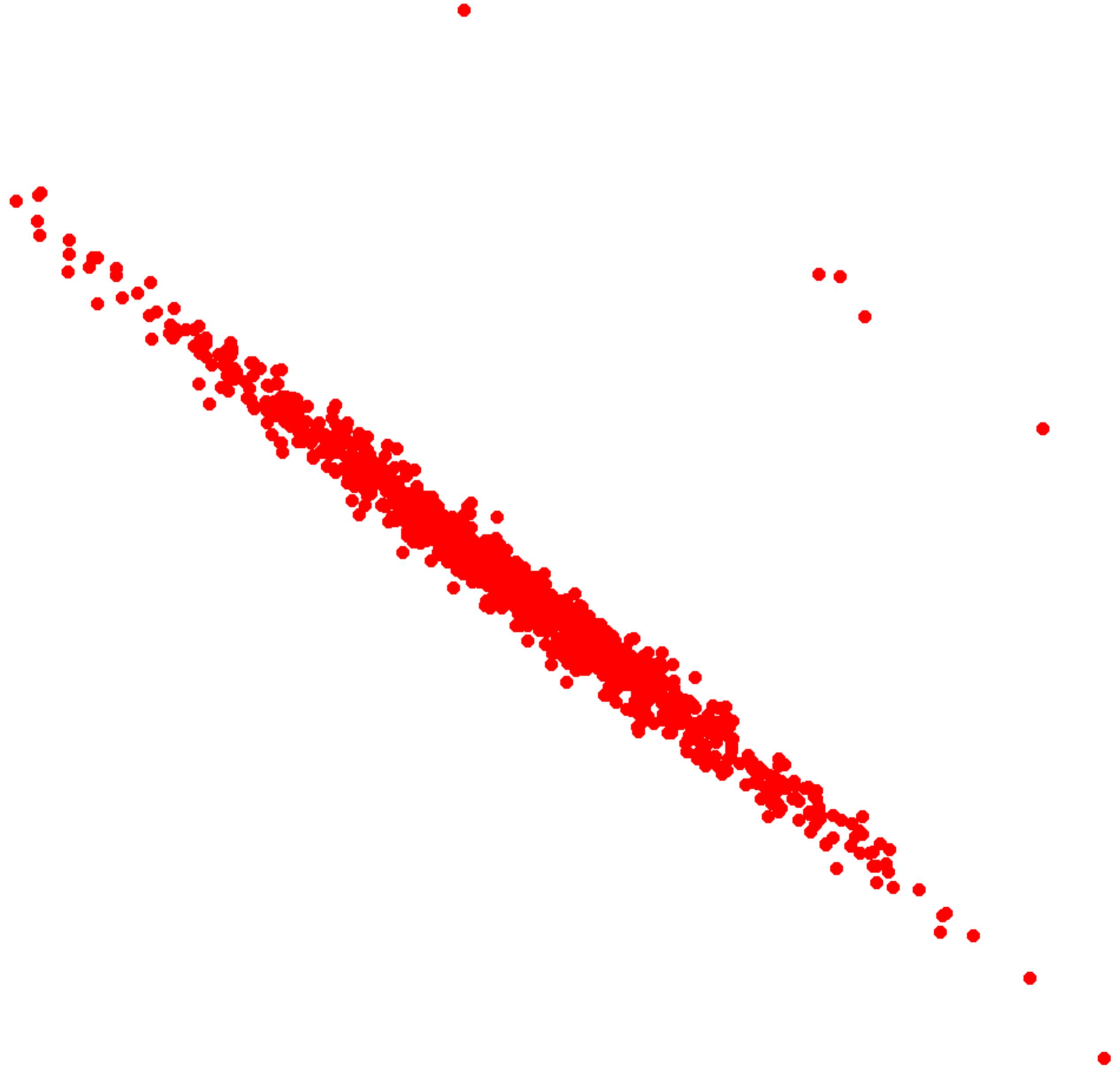
-0.92346



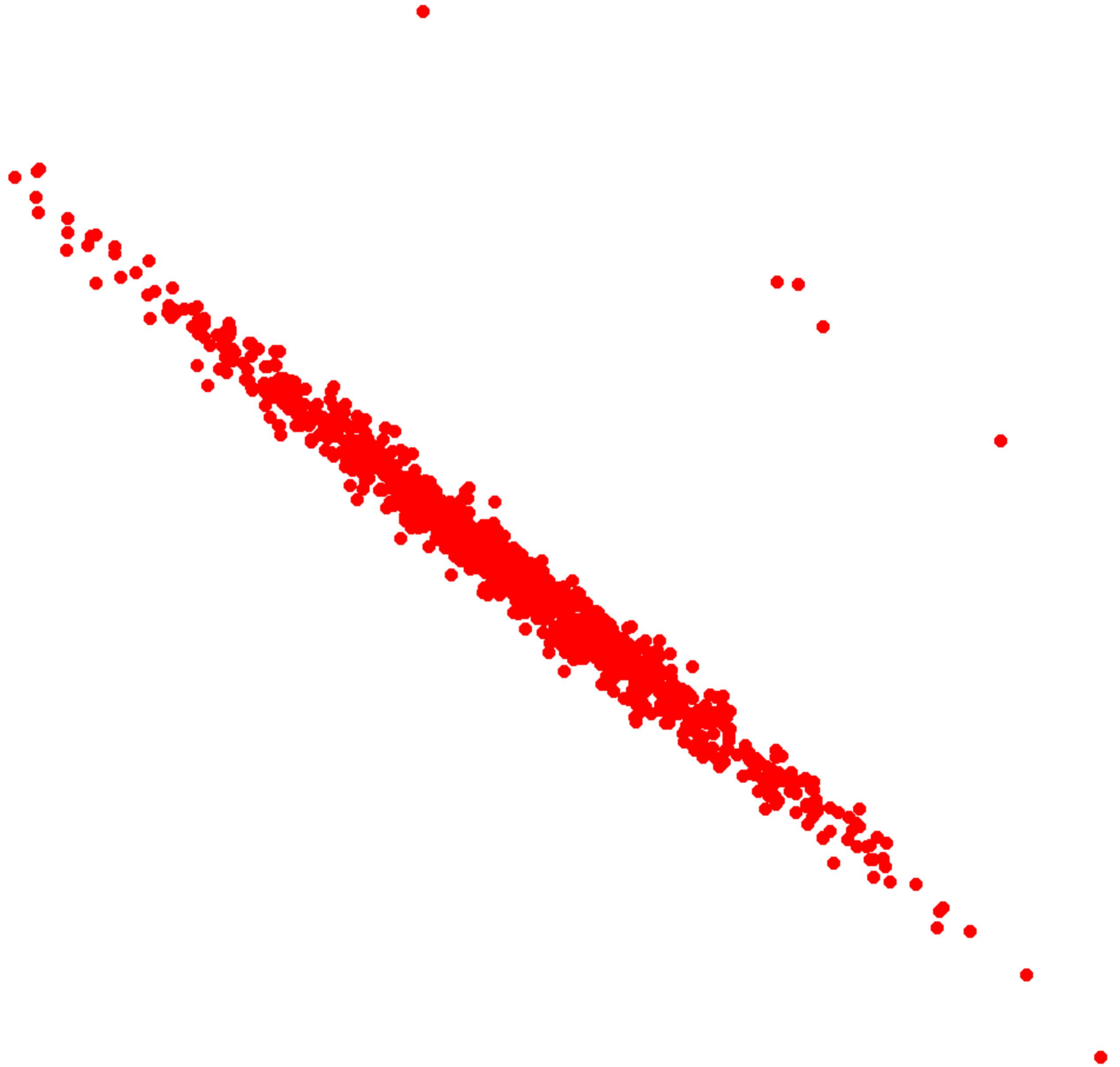
-0.93536



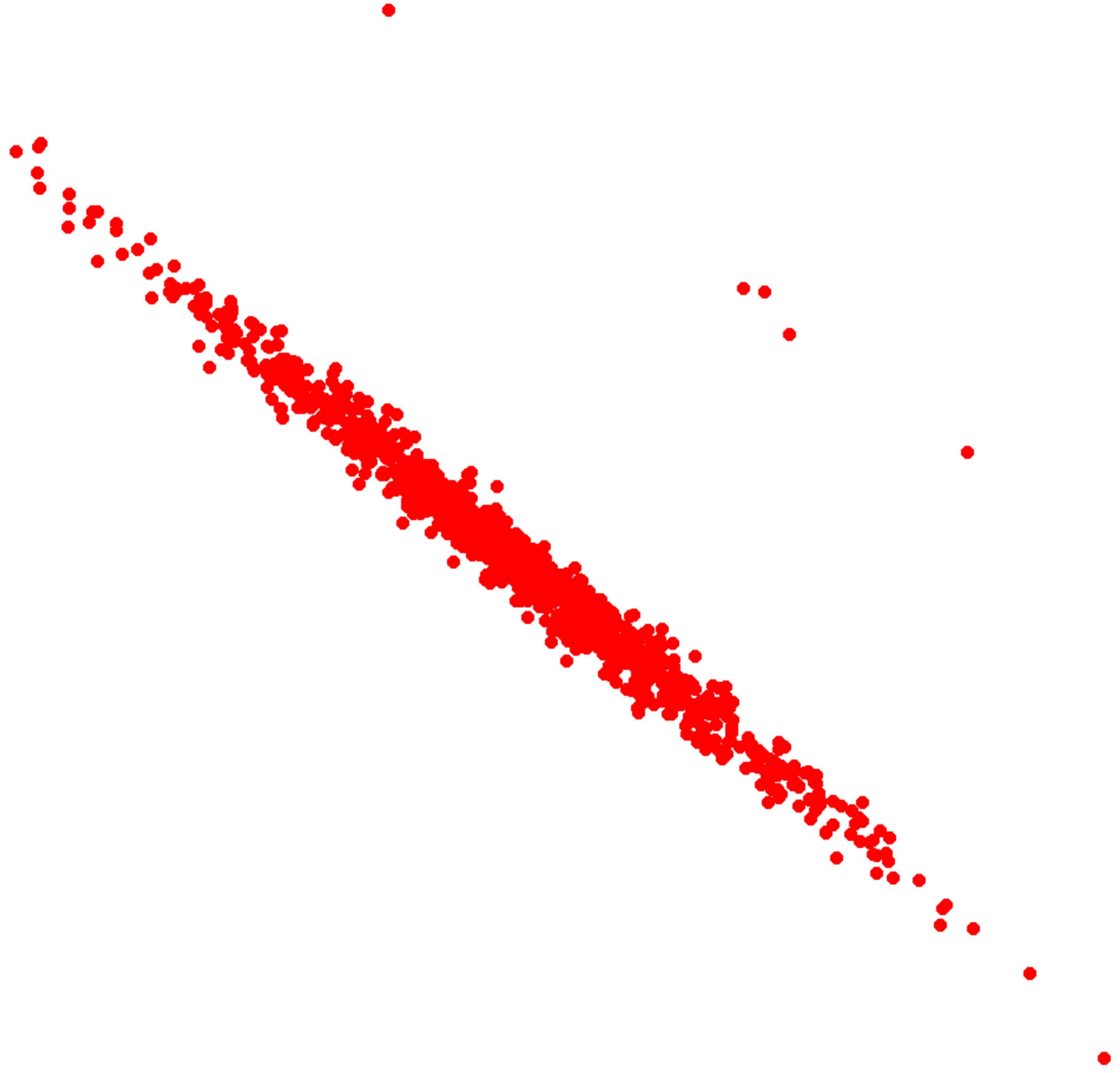
-0.94567



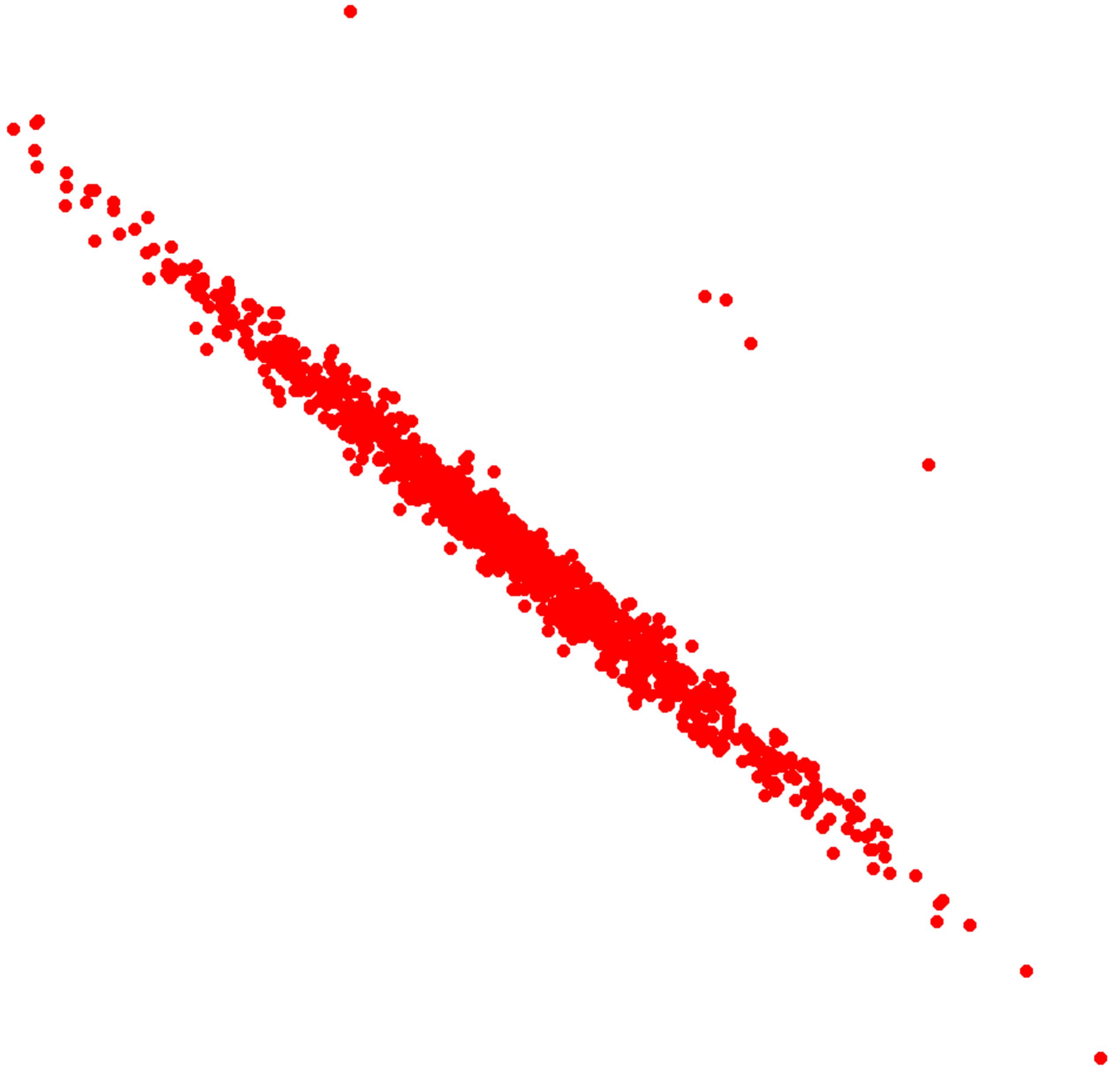
-0.9545



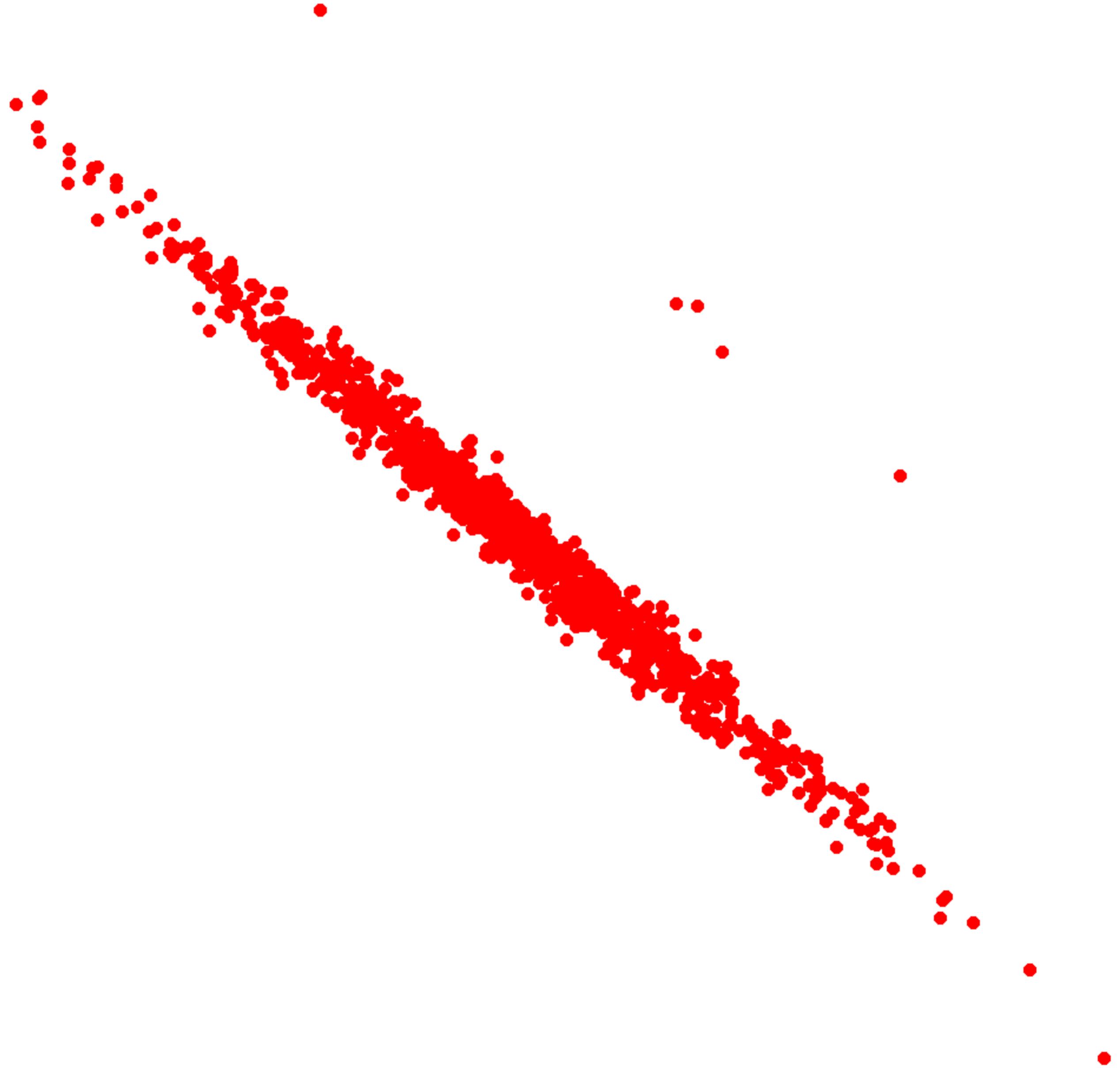
-0.96198



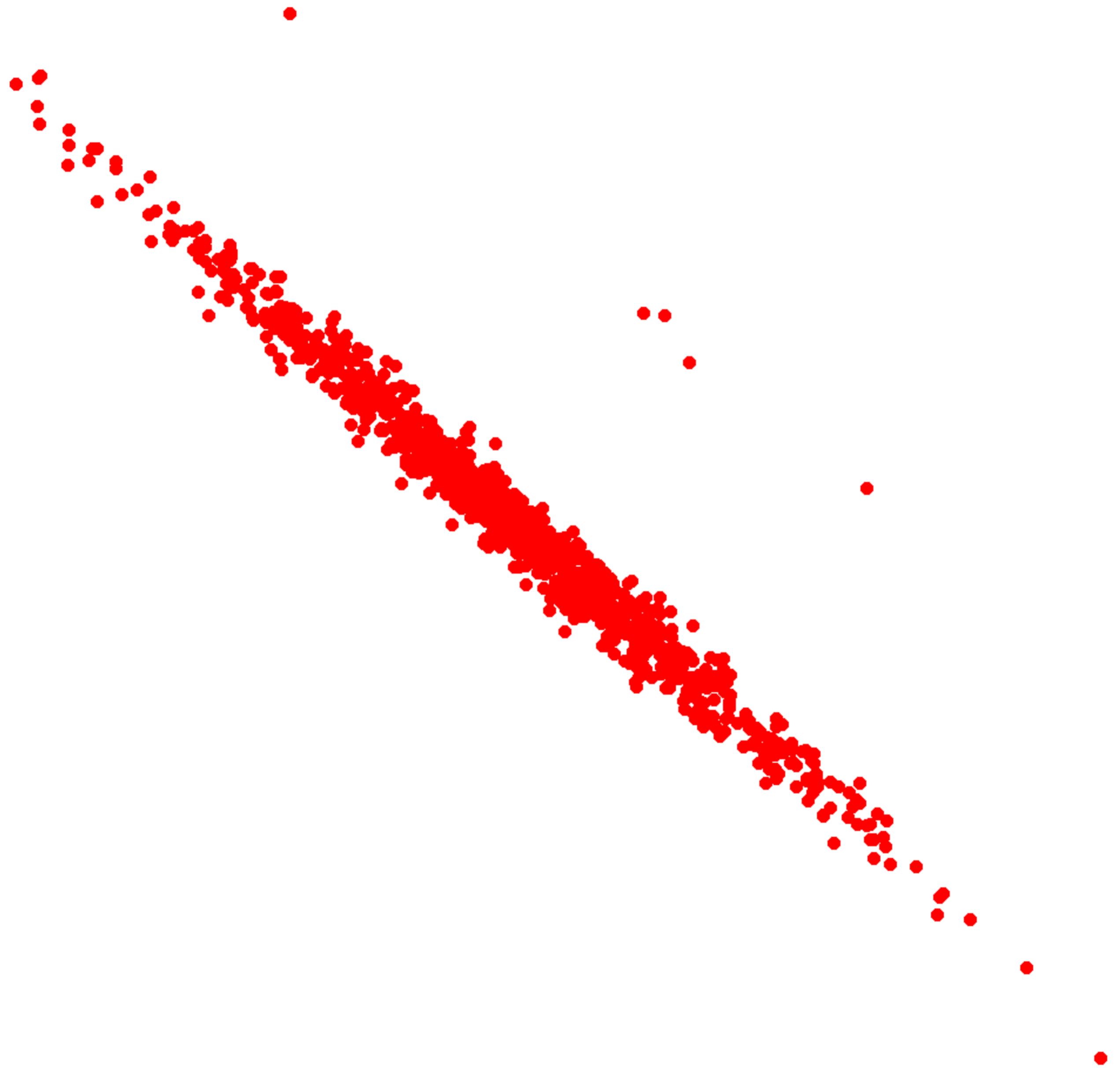
-0.96825



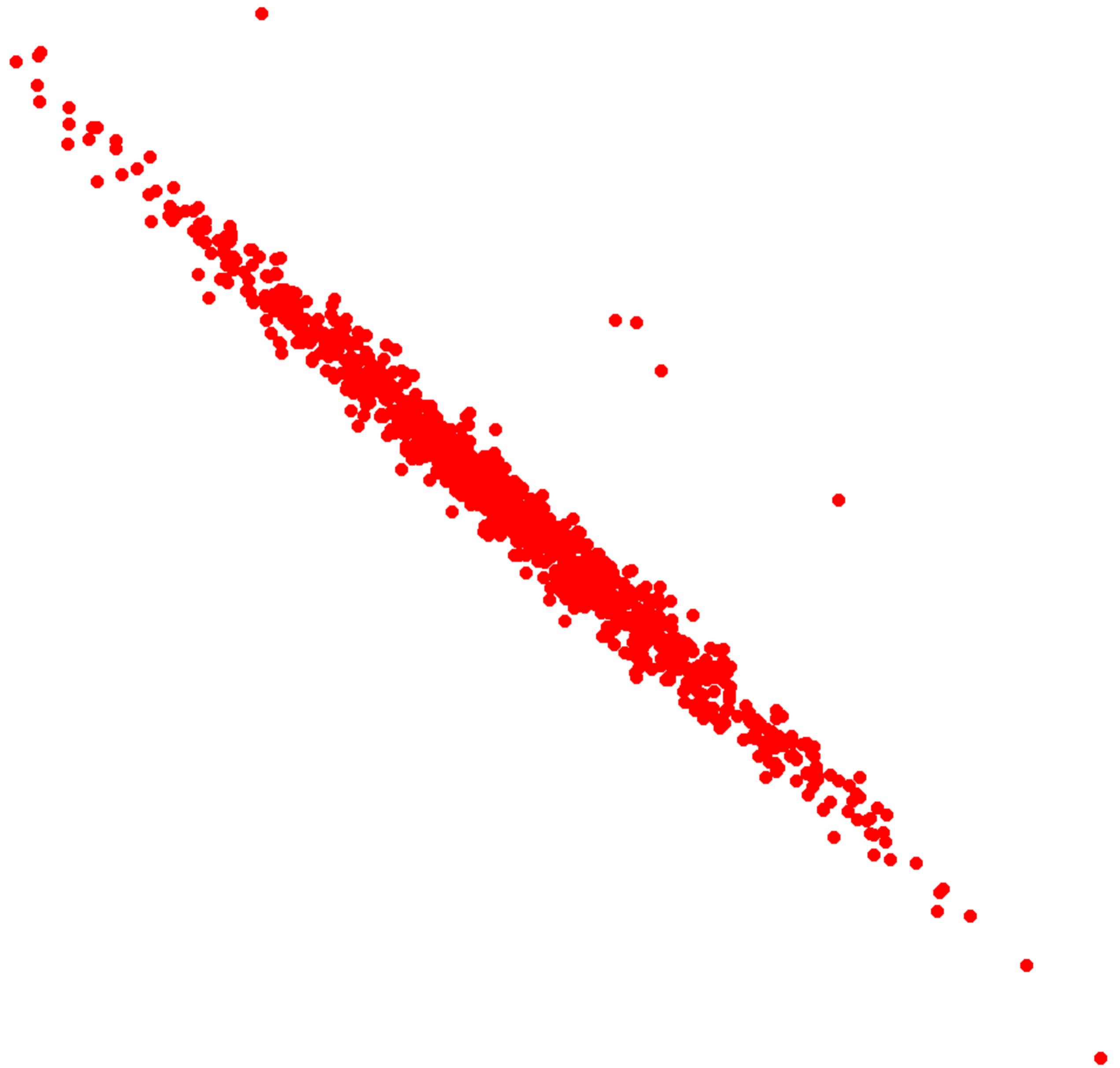
-0.97343



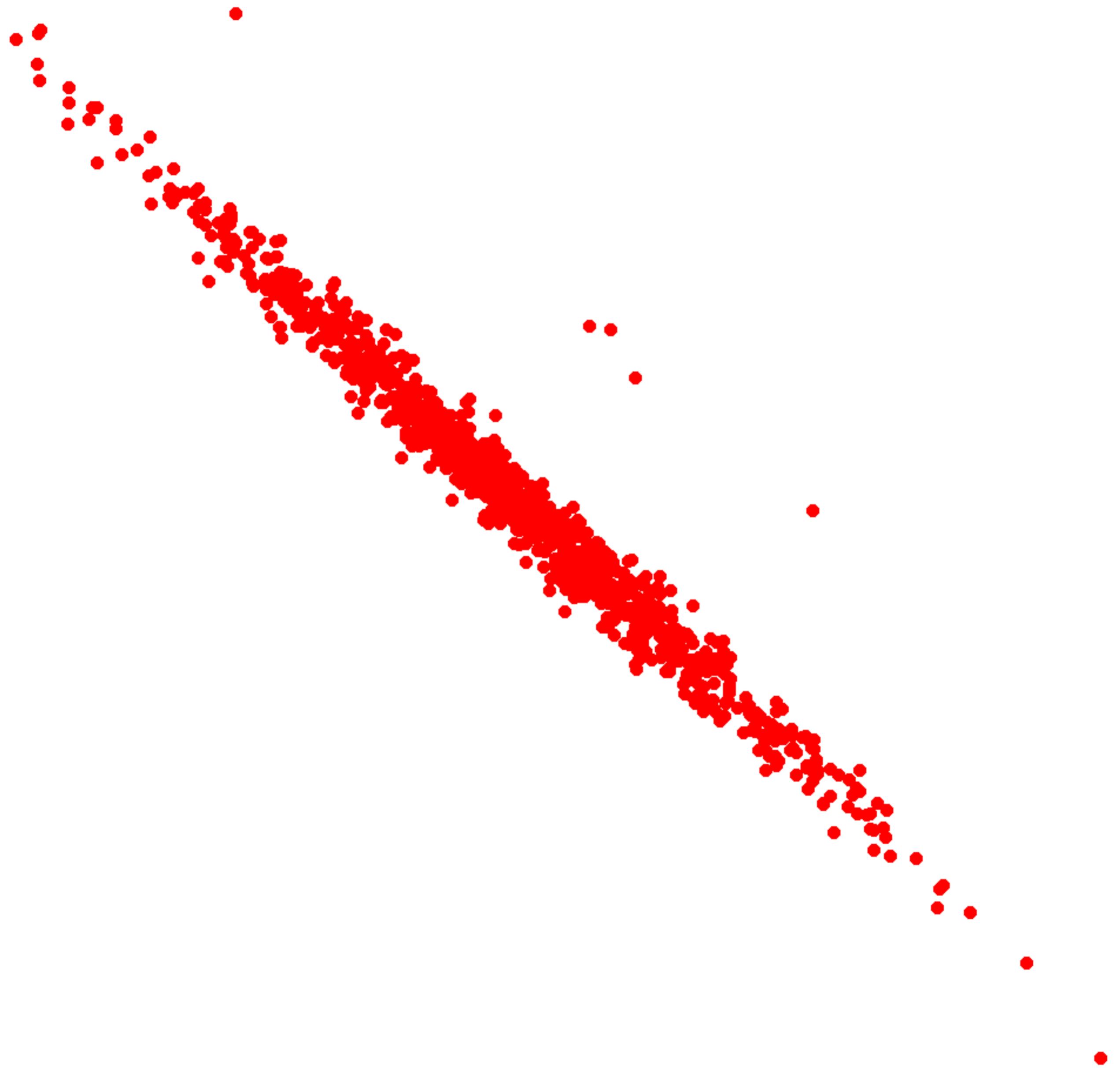
-0.97764



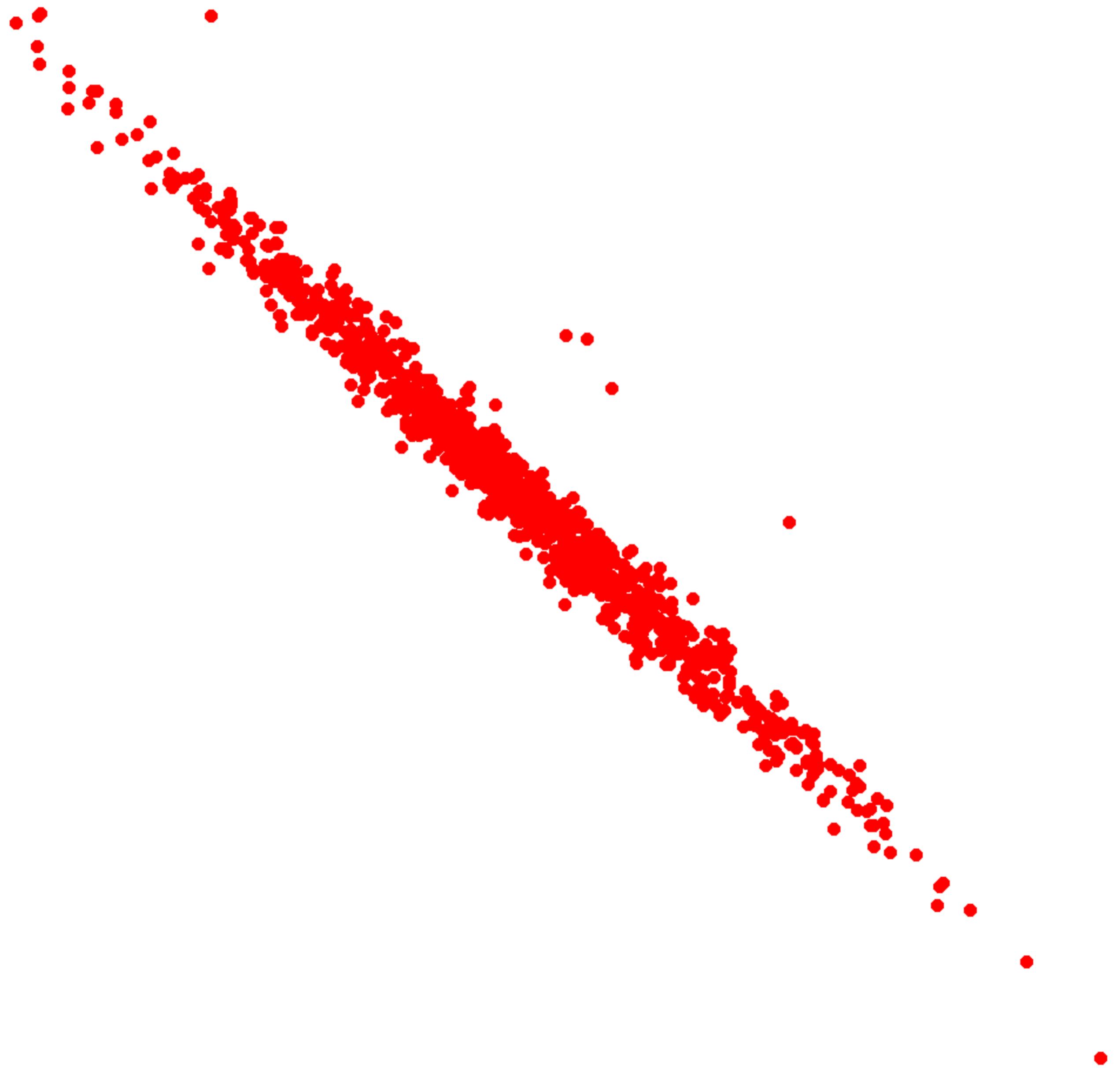
-0.98101



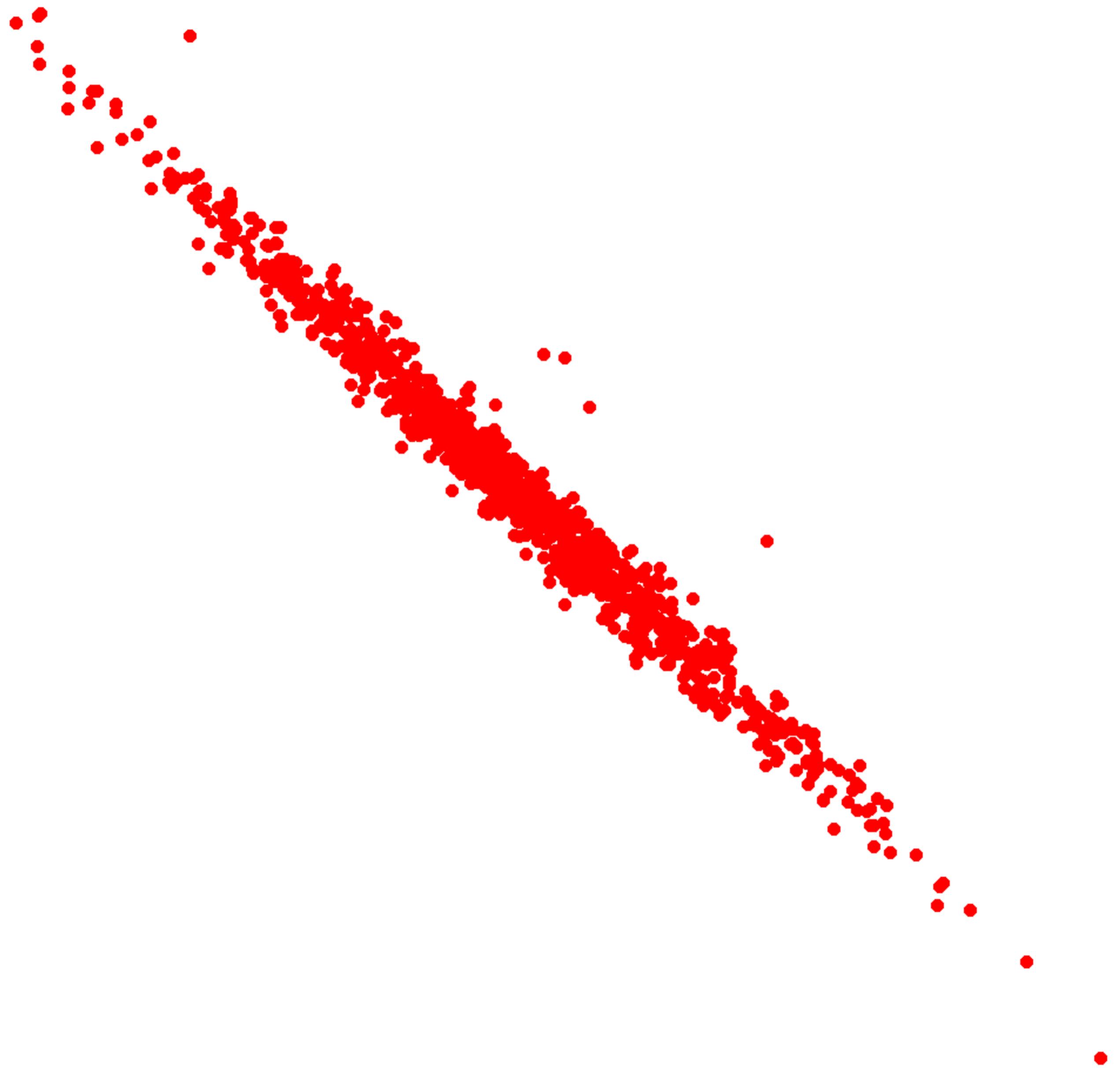
-0.98365



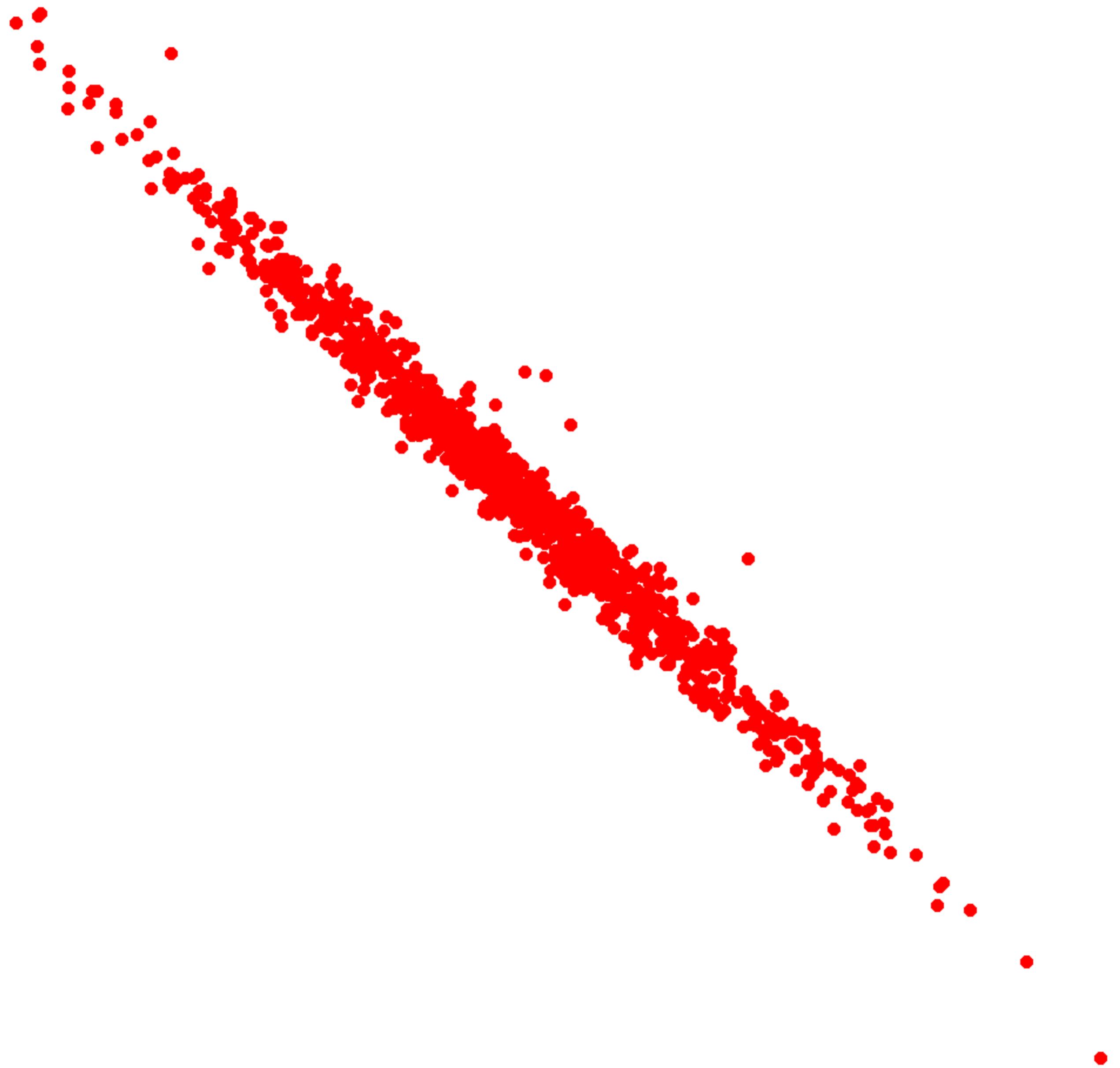
-0.98567



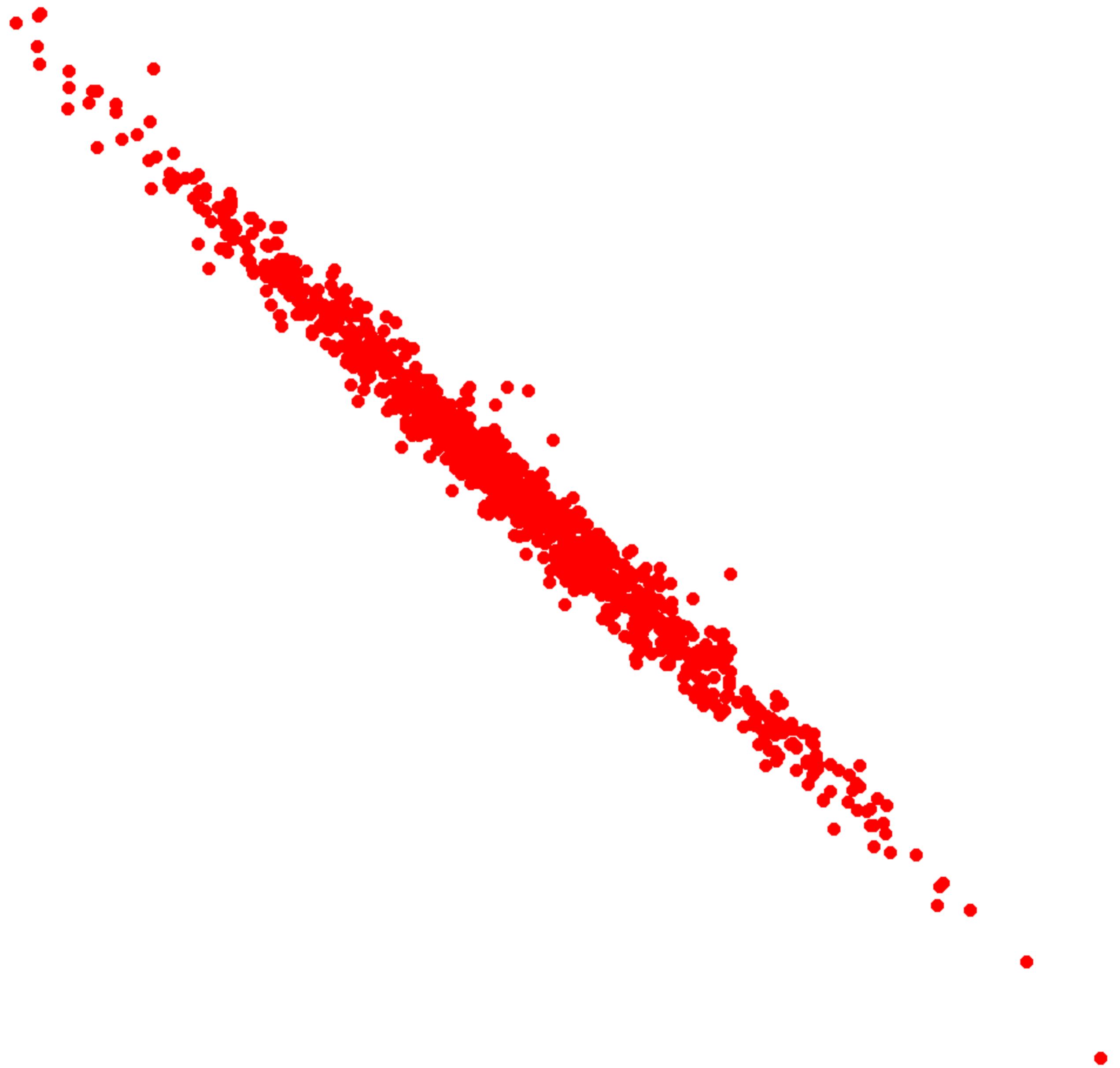
-0.98717



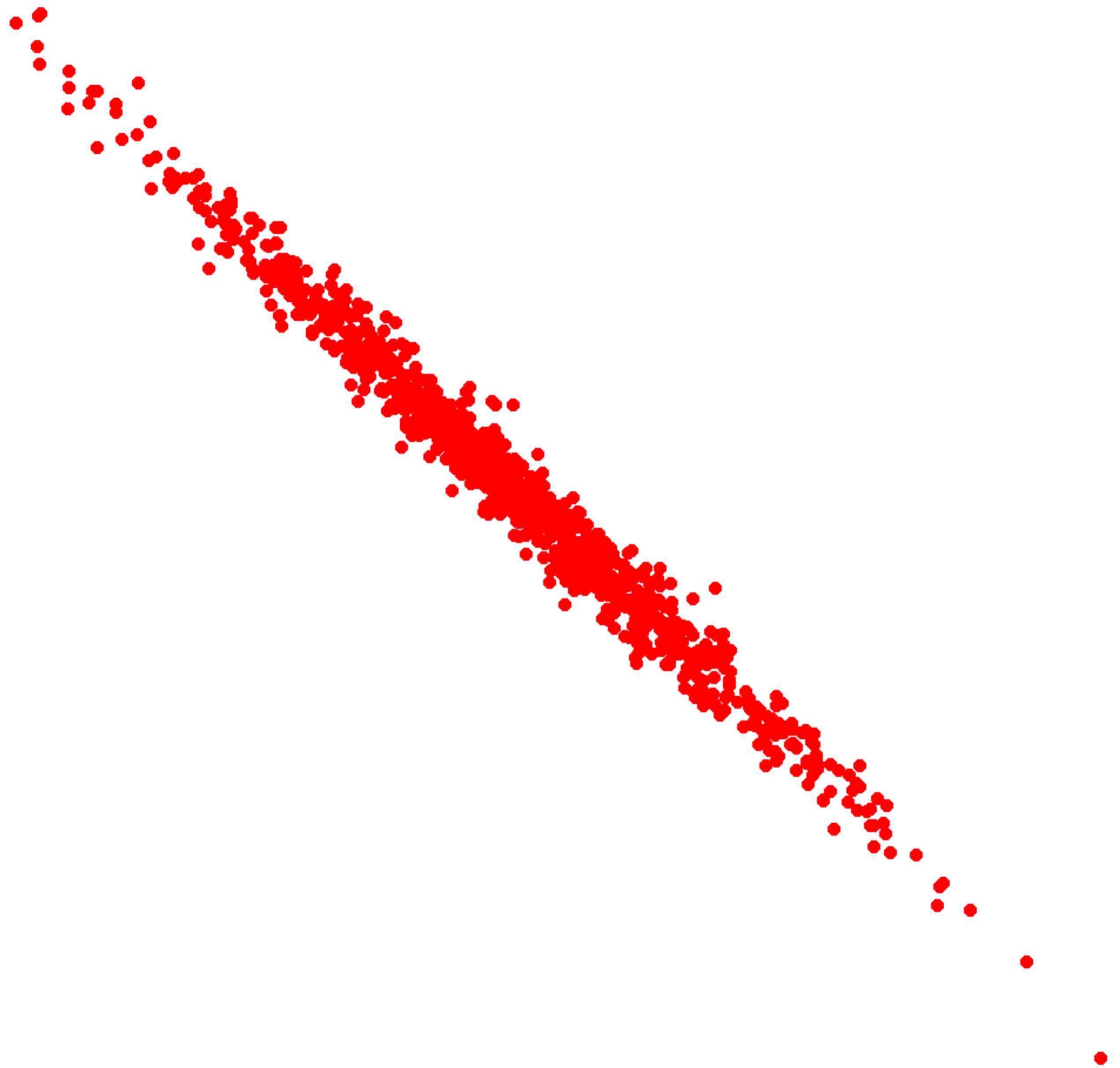
-0.98826



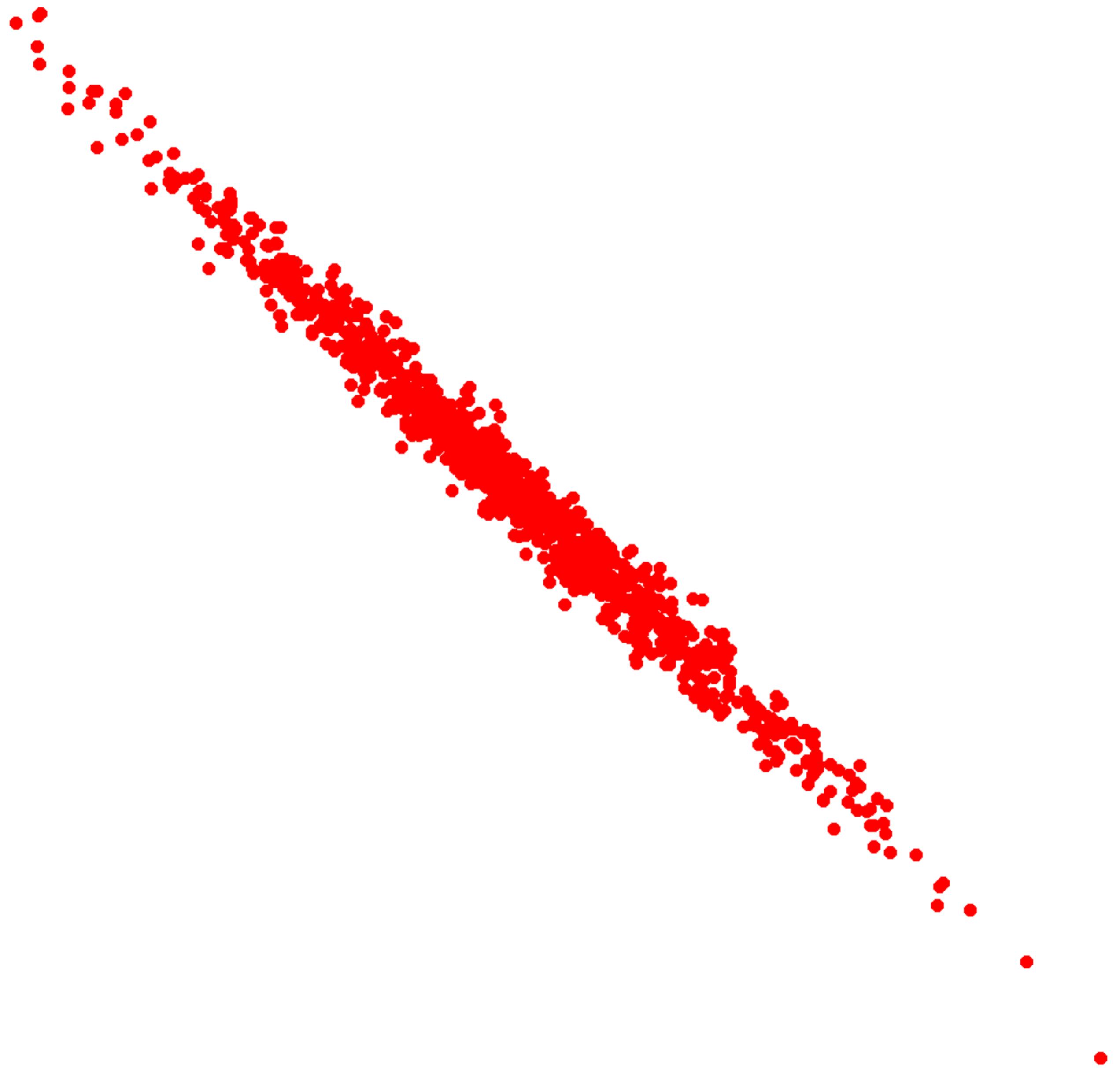
-0.989



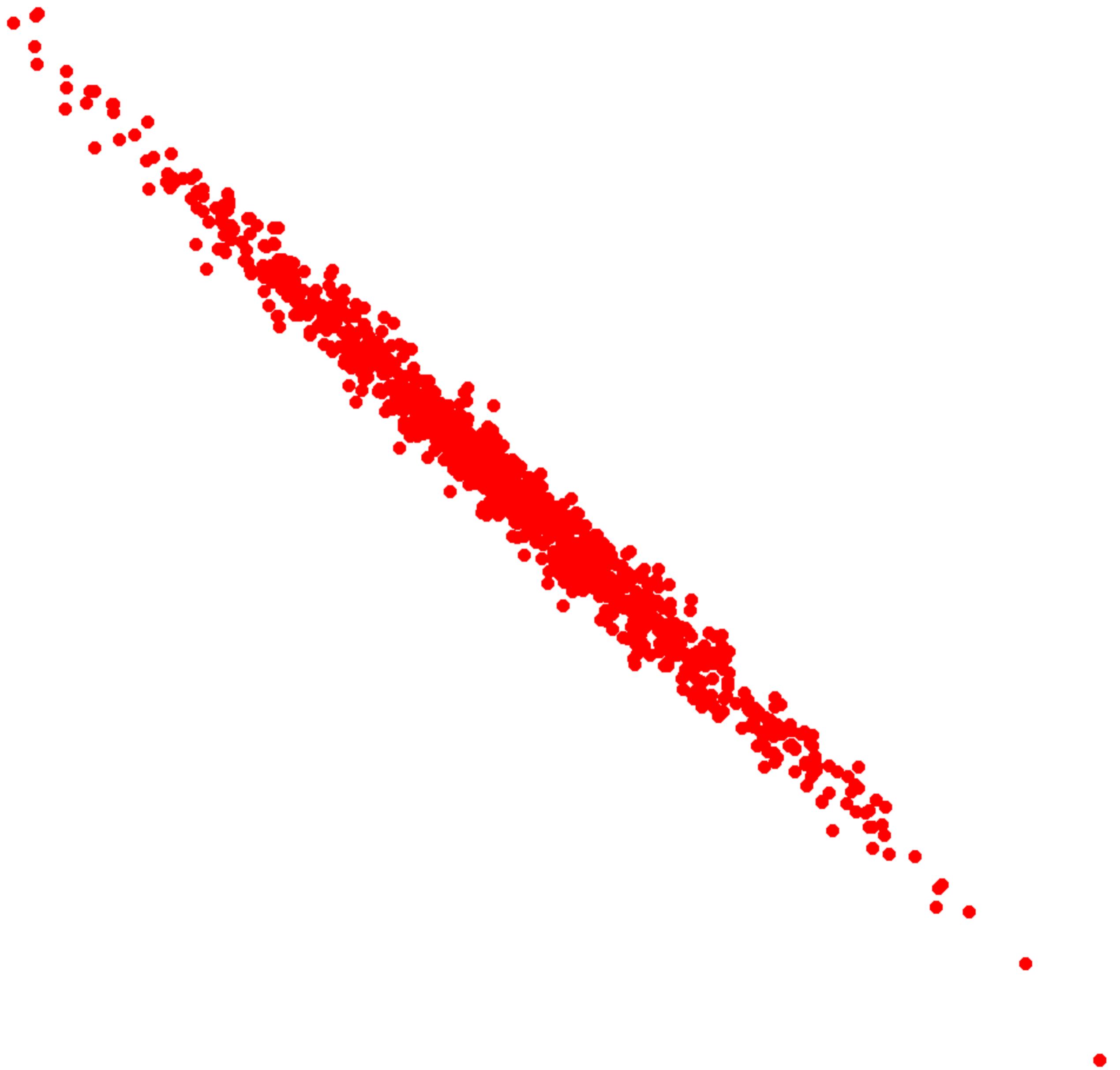
-0.98949



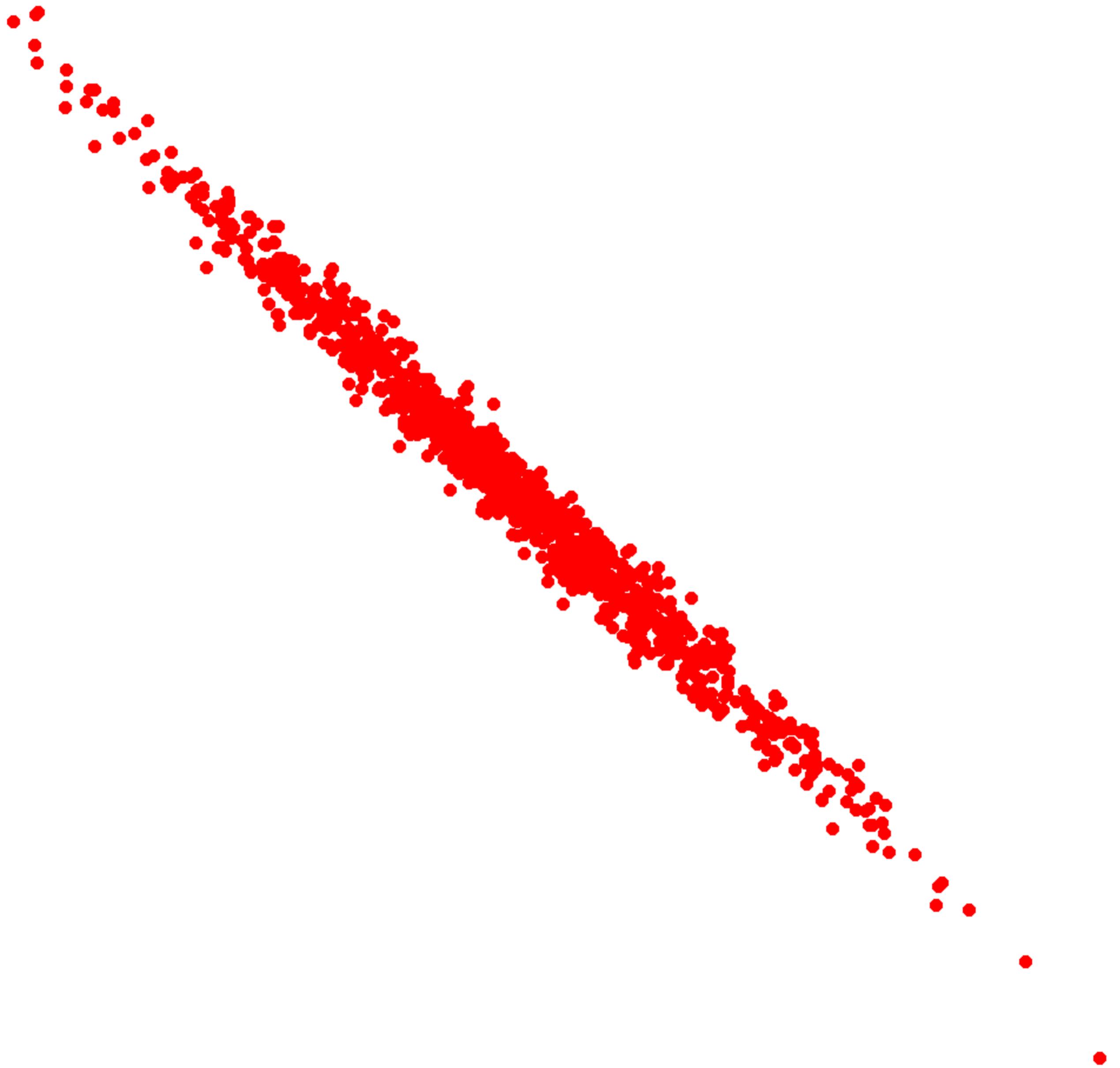
-0.98978



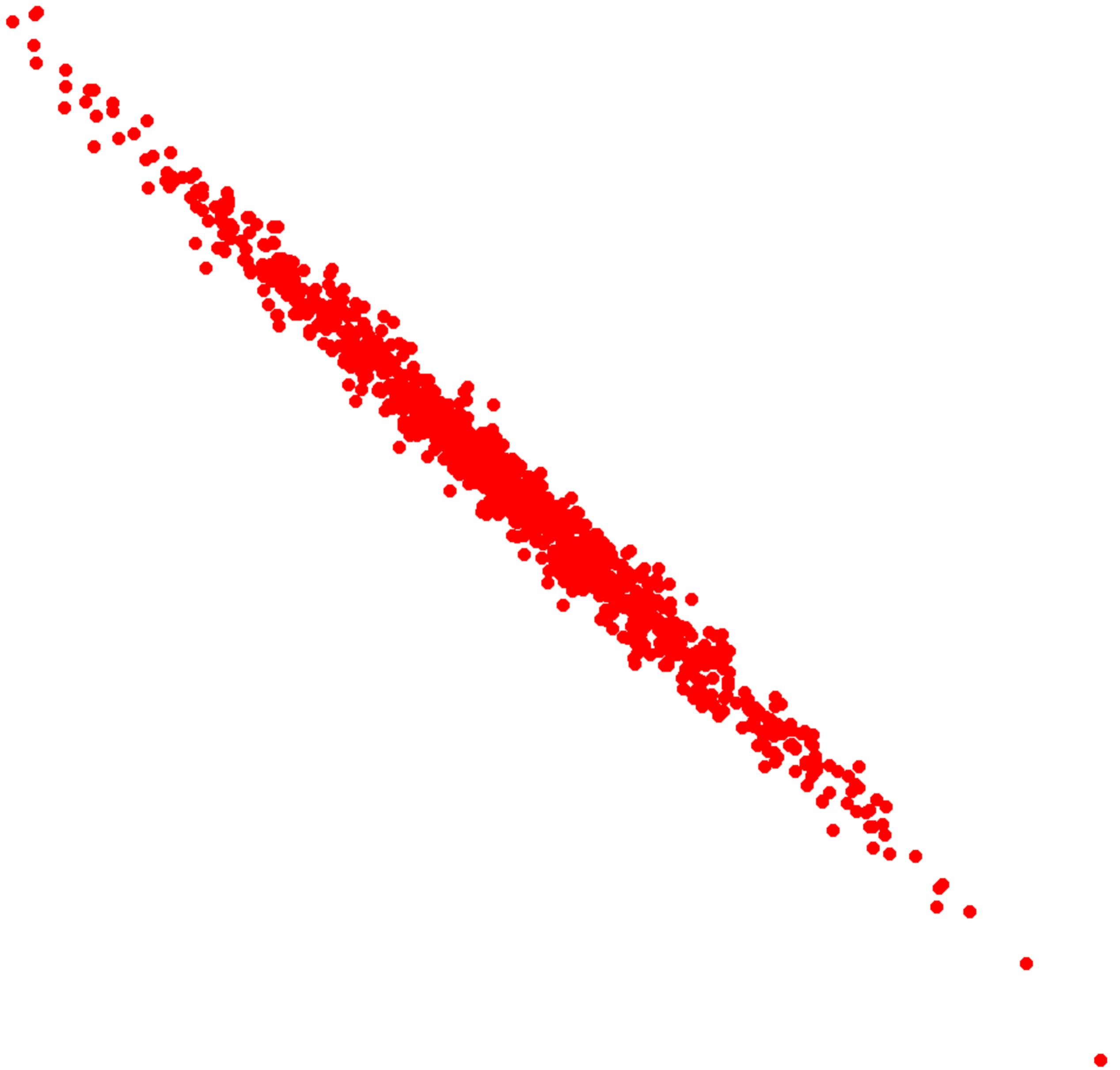
-0.98993



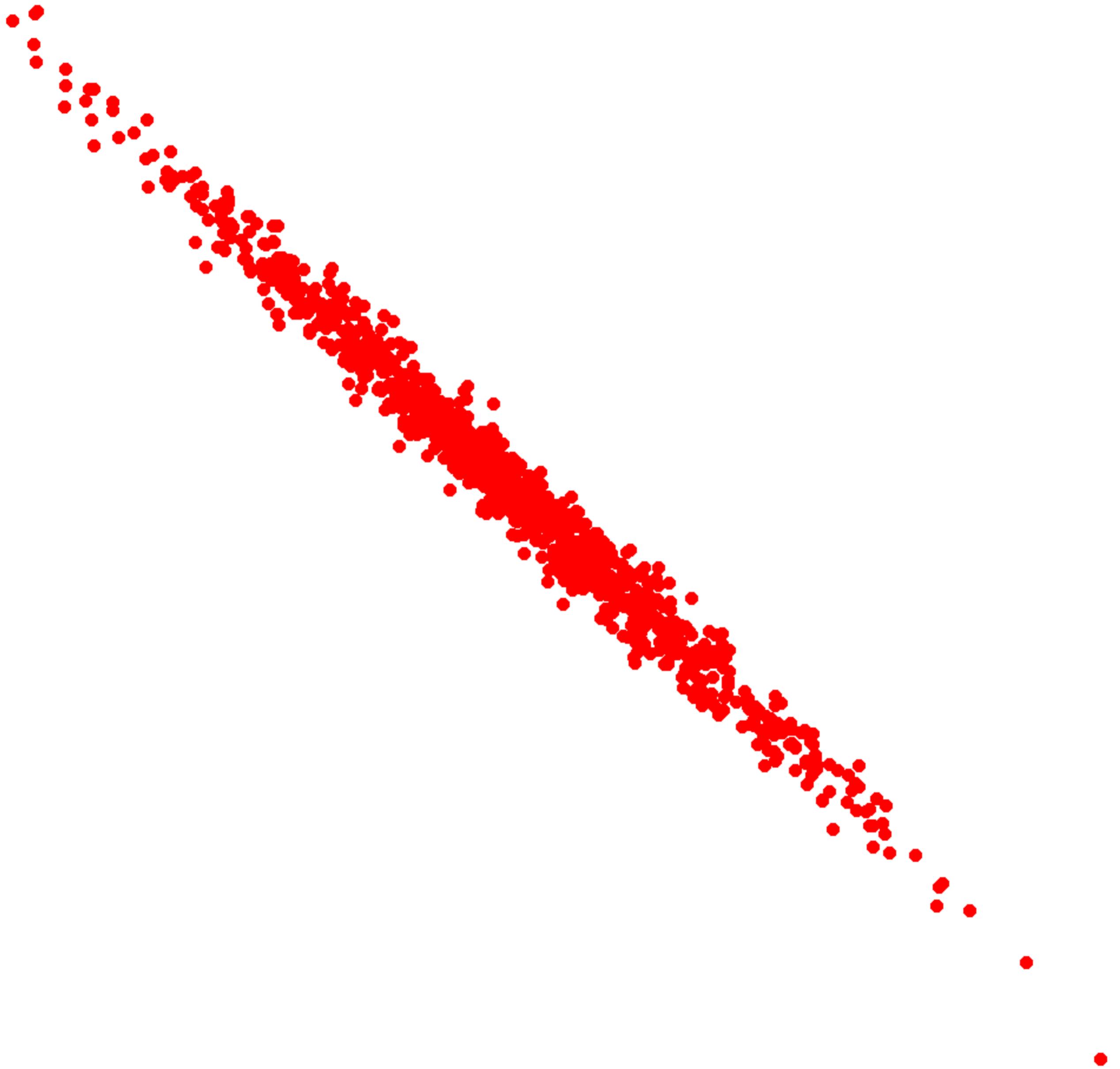
-0.99



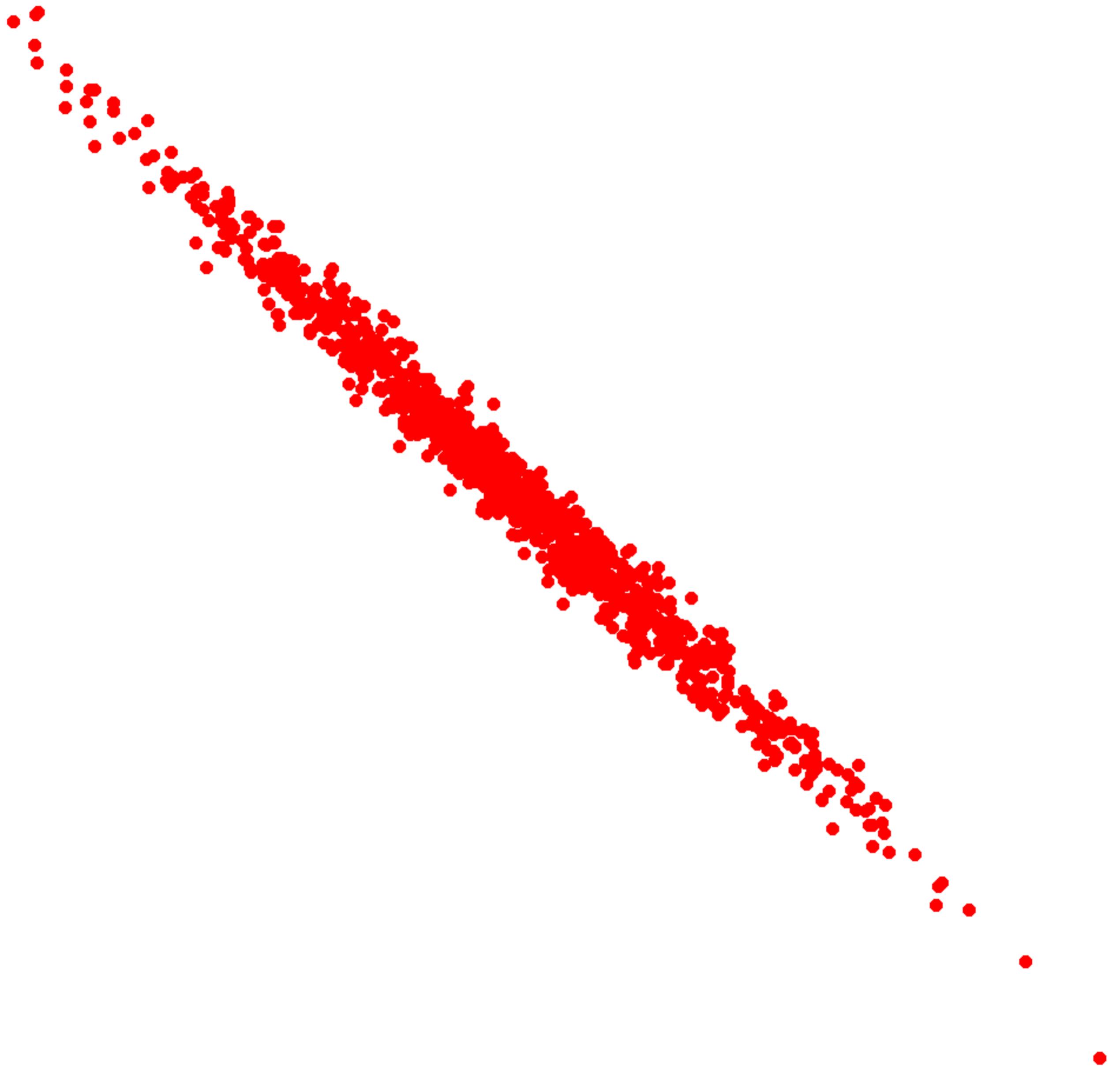
-0.99001



-0.99001



-0.99



КОРРЕЛЯЦИЯ СПИРМЕНА

КОРРЕЛЯЦИЯ СПИРМЕНА

- » Коэффициент корреляции Спирмена — мера силы монотонной взаимосвязи
- » Равен коэффициенту корреляции Пирсона между рангами наблюдений

$$\rho_{X_1 X_2} \in [-1, 1]$$

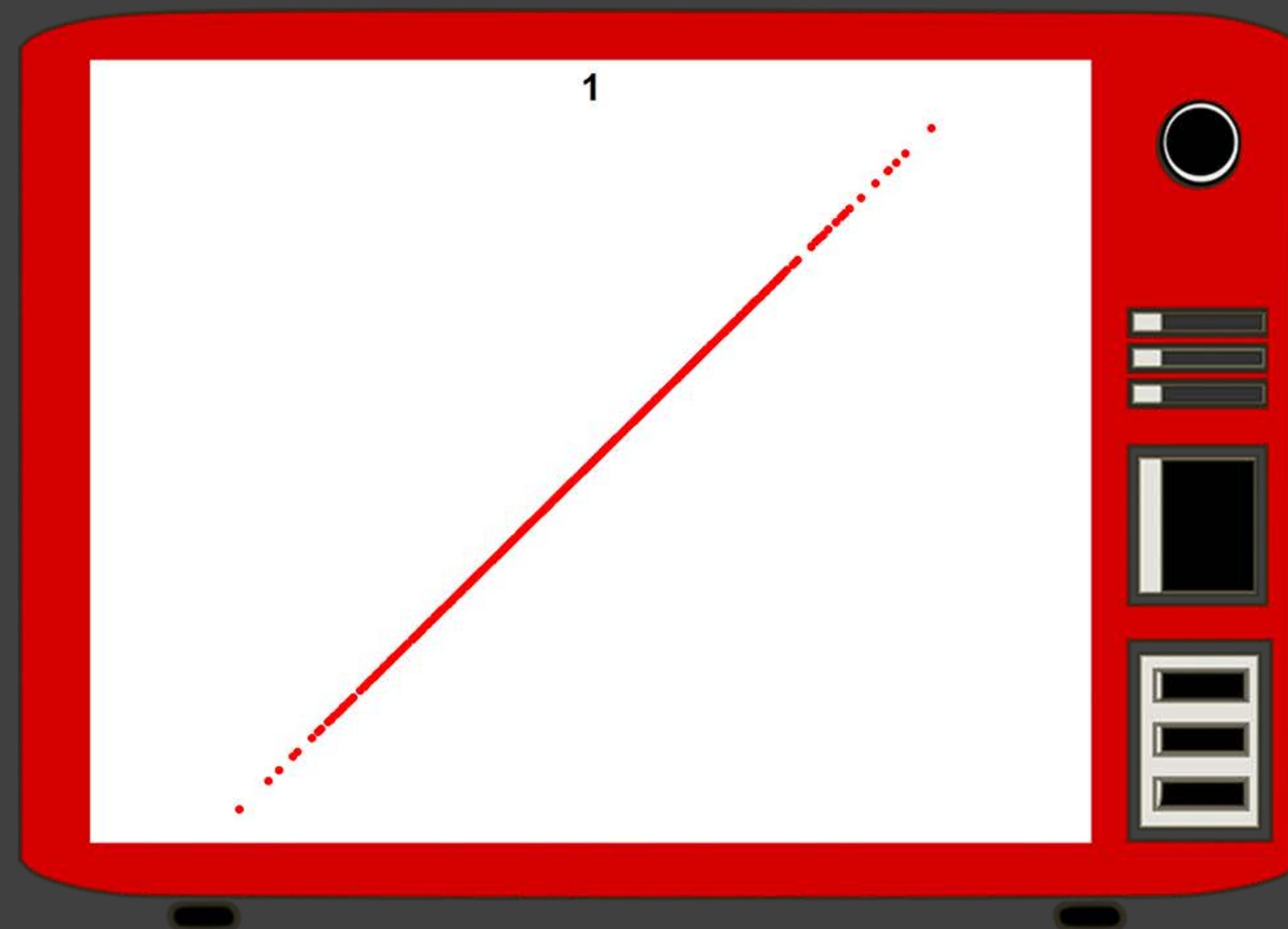
ВЫБОРОЧНЫЙ КОЭФФИЦИЕНТ

- » Выборка пар (X_{1i}, X_{2i}) , $i = 1, \dots, n$;
 $\text{rank}(X_{1i}), \text{rank}(X_{2i})$ — ранги
- » Выборочный коэффициент корреляции Спирмена:

$$\rho_{X_1 X_2} = \frac{\sum_{i=1}^n \left(\text{rank}(X_{1i}) - \frac{n+1}{2} \right) \left(\text{rank}(X_{2i}) - \frac{n+1}{2} \right)}{\frac{1}{12} (n^3 - n)} = \\ = 1 - \frac{6}{n^3 - n} \sum_{i=1}^n (\text{rank}(X_{1i}) - \text{rank}(X_{2i}))^2$$

СЖИМАЕМ ОБЛАКО





СГИБАЕМ ОБЛАКО



СЛОЖНЫЕ ЗАВИСИМОСТИ



ВЫБРОСЫ

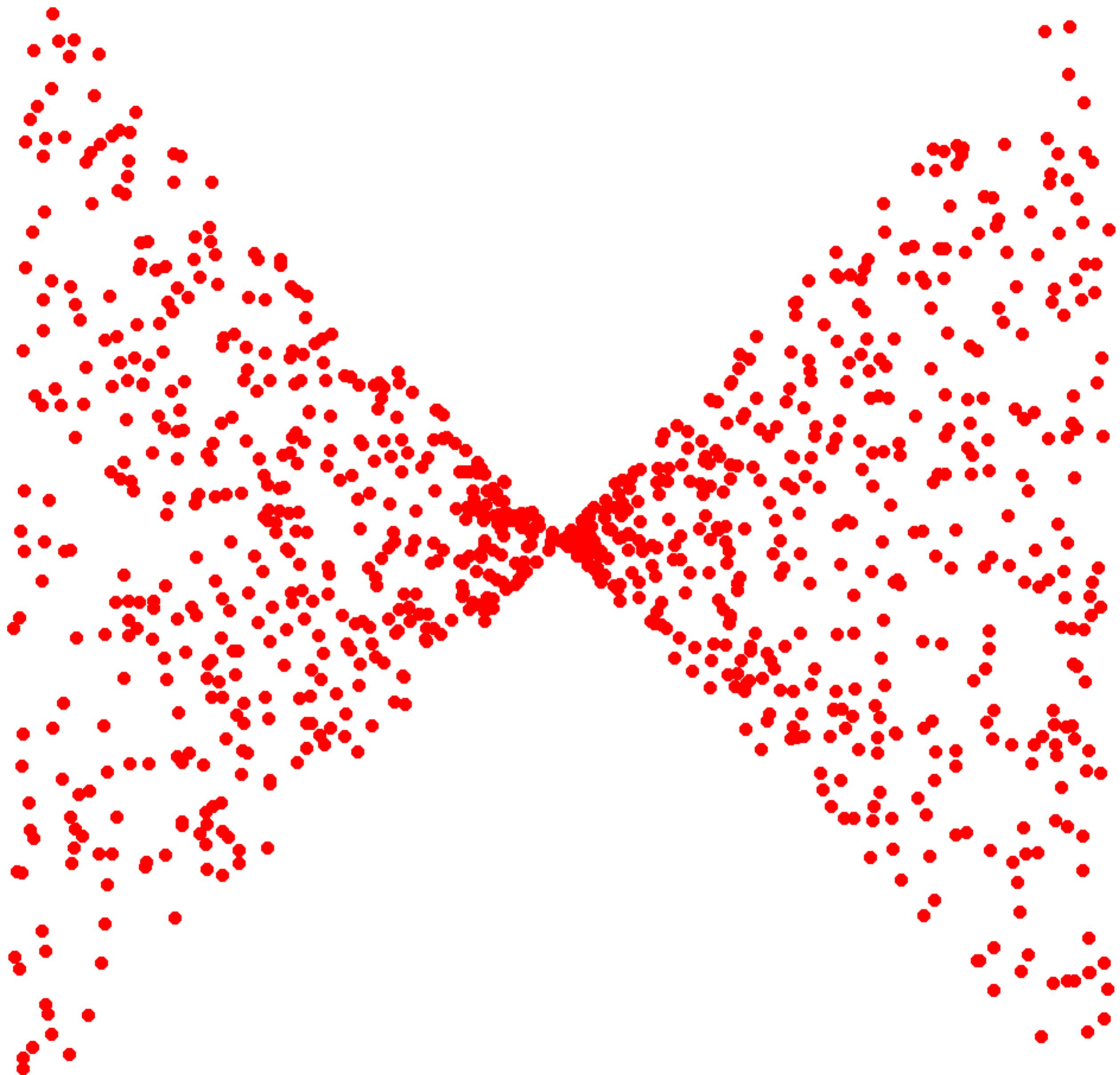


- » Коэффициент корреляции Спирмена — мера силы монотонной взаимосвязи
- » Далее: корреляция Мэтьюса и коэффициент Крамера

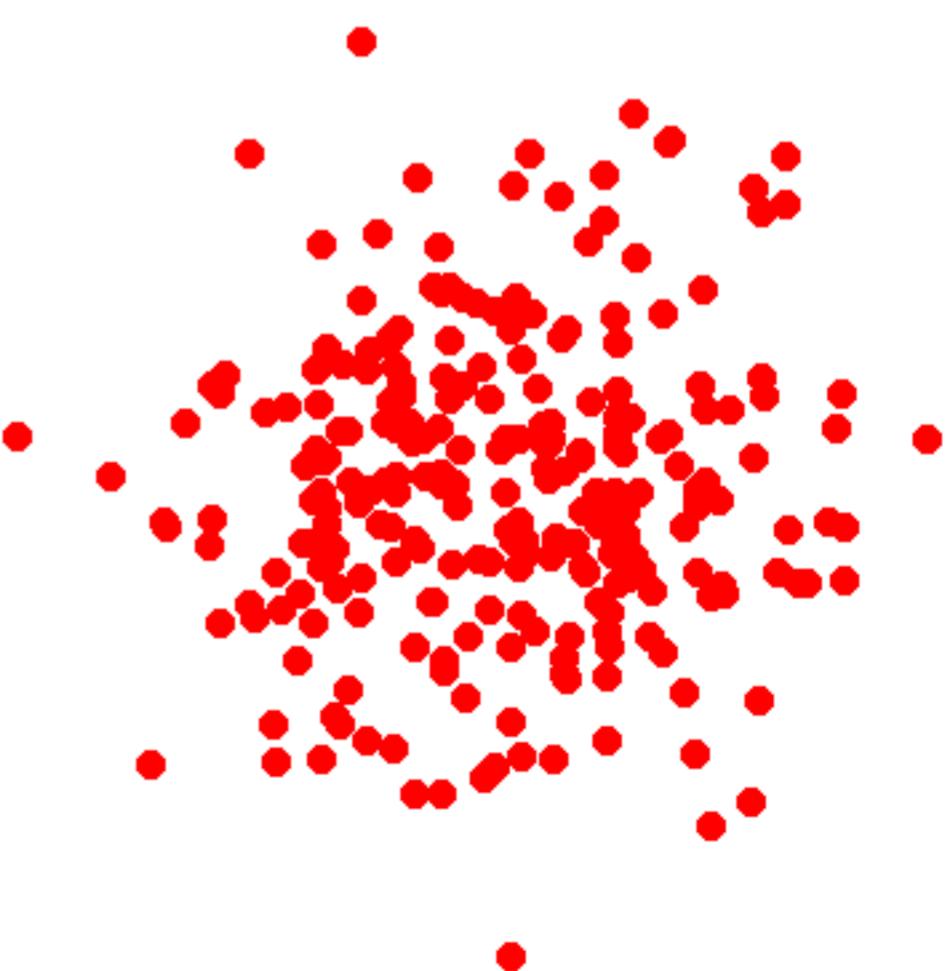
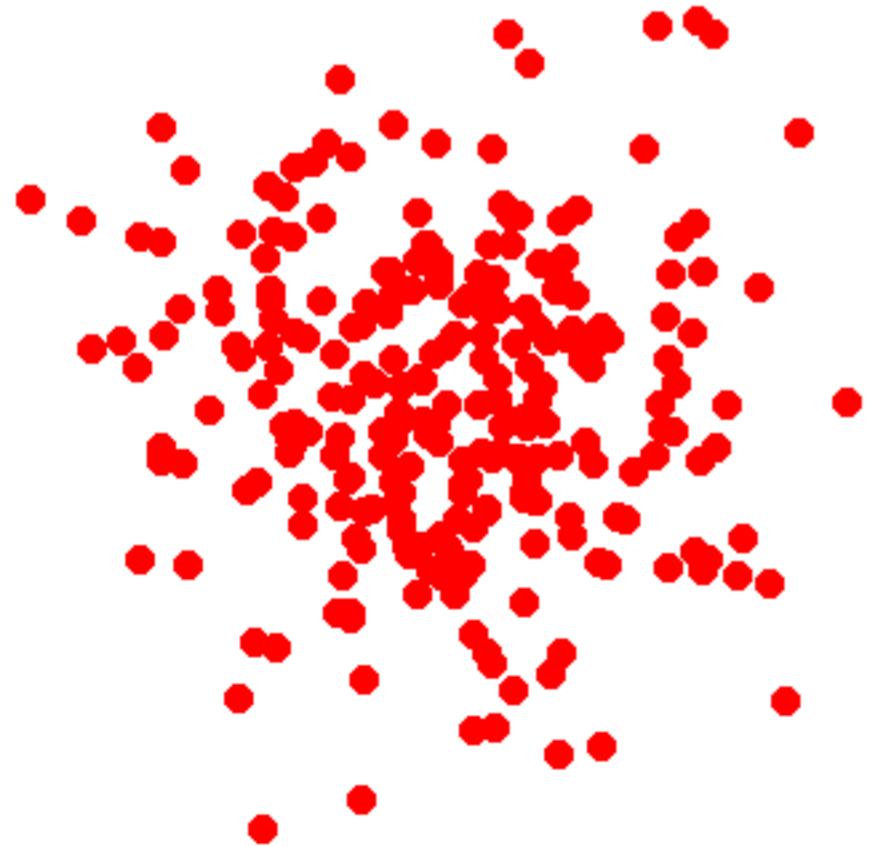
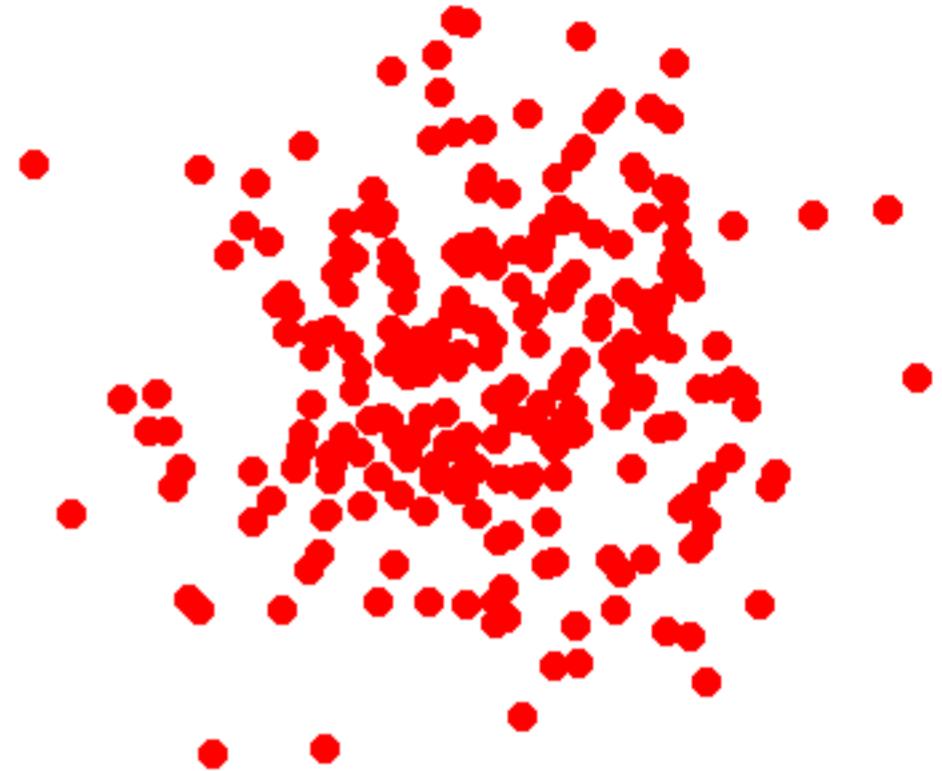
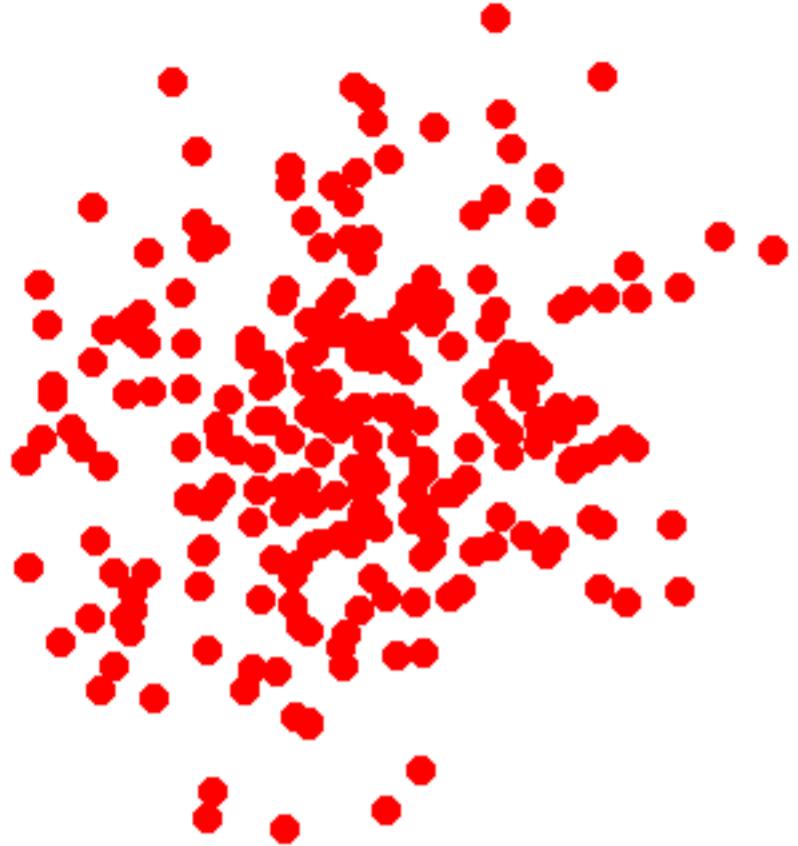
0.01869



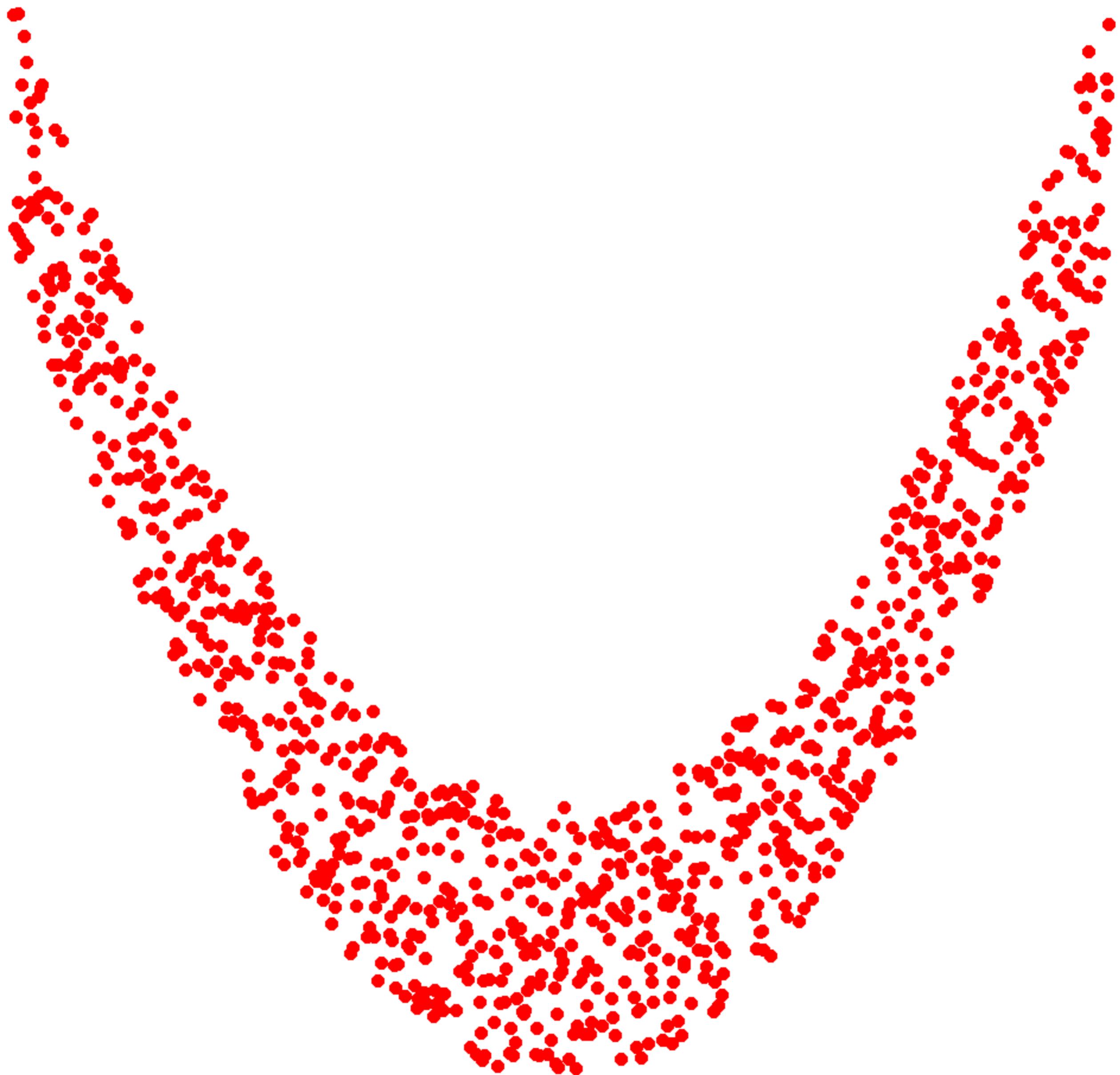
-0.01245



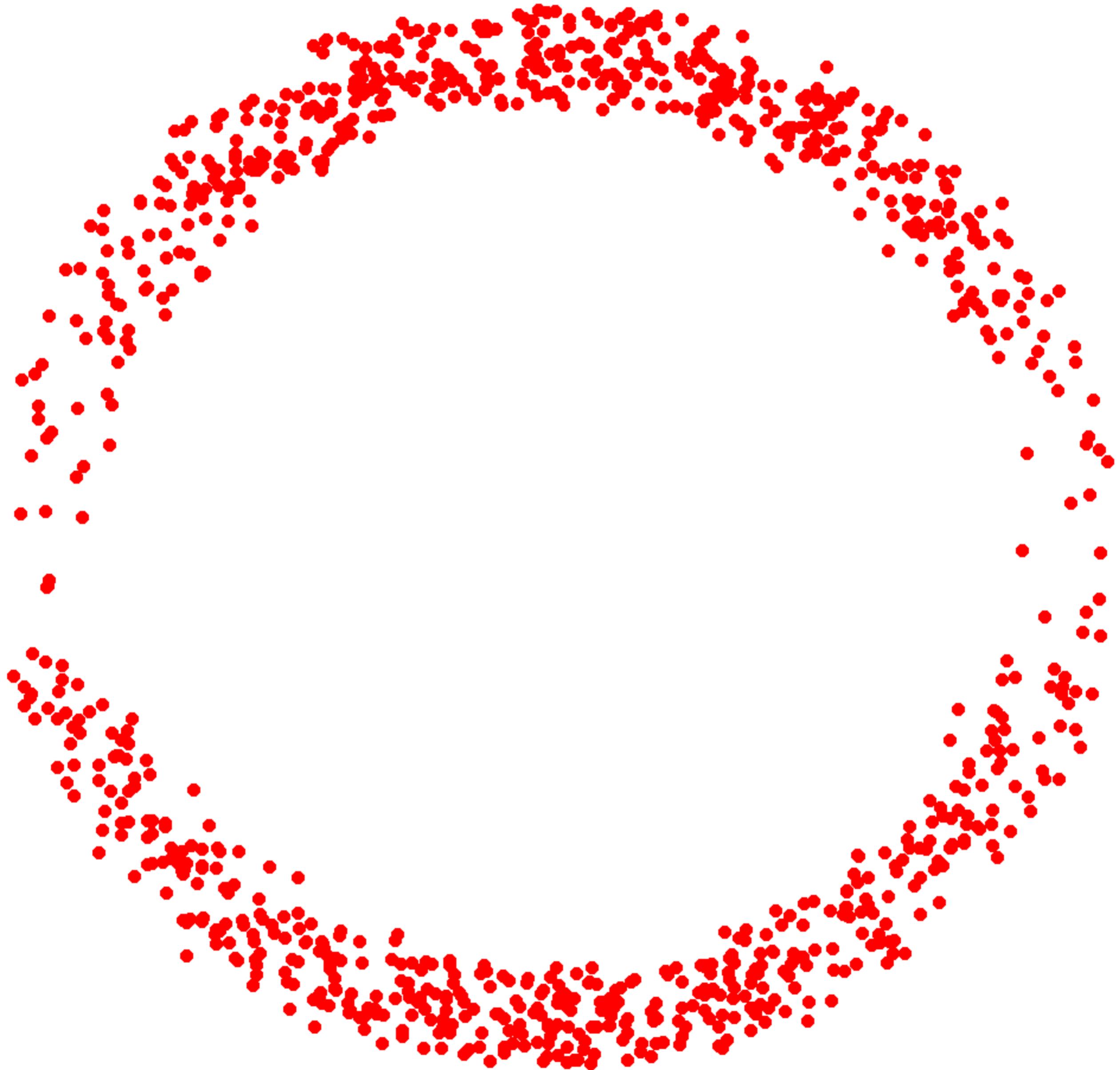
0.03164



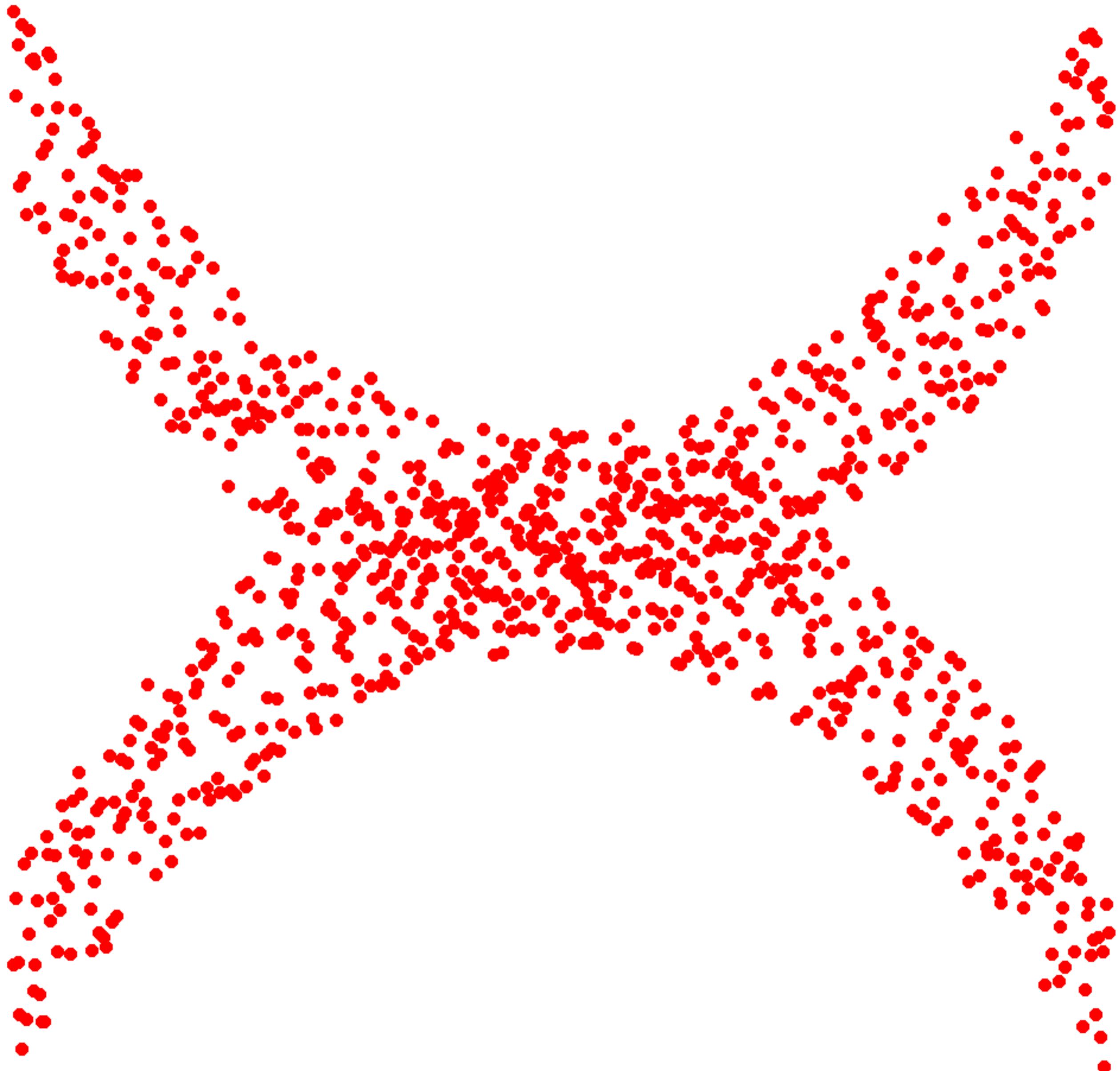
-0.00309



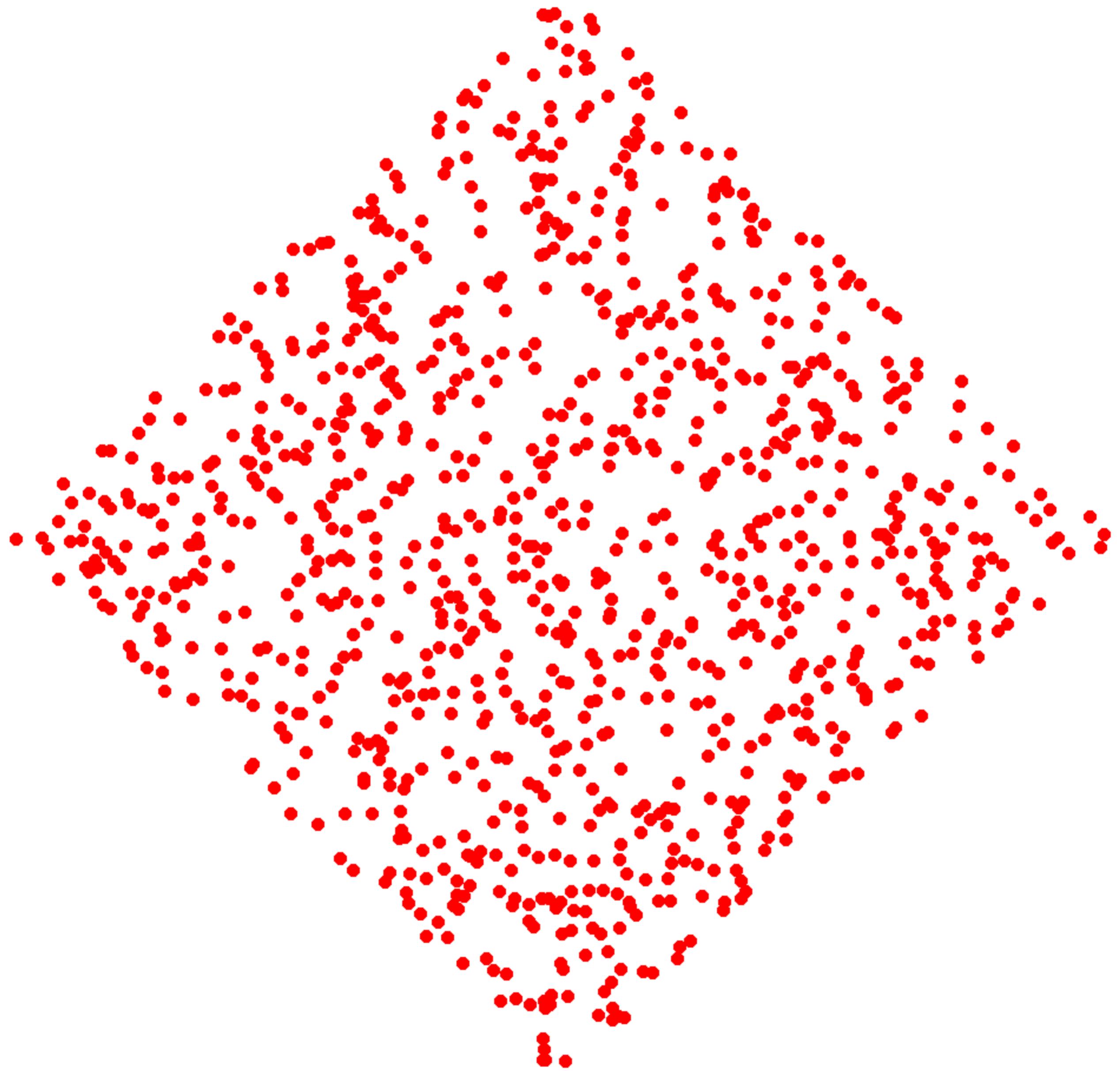
0.00785



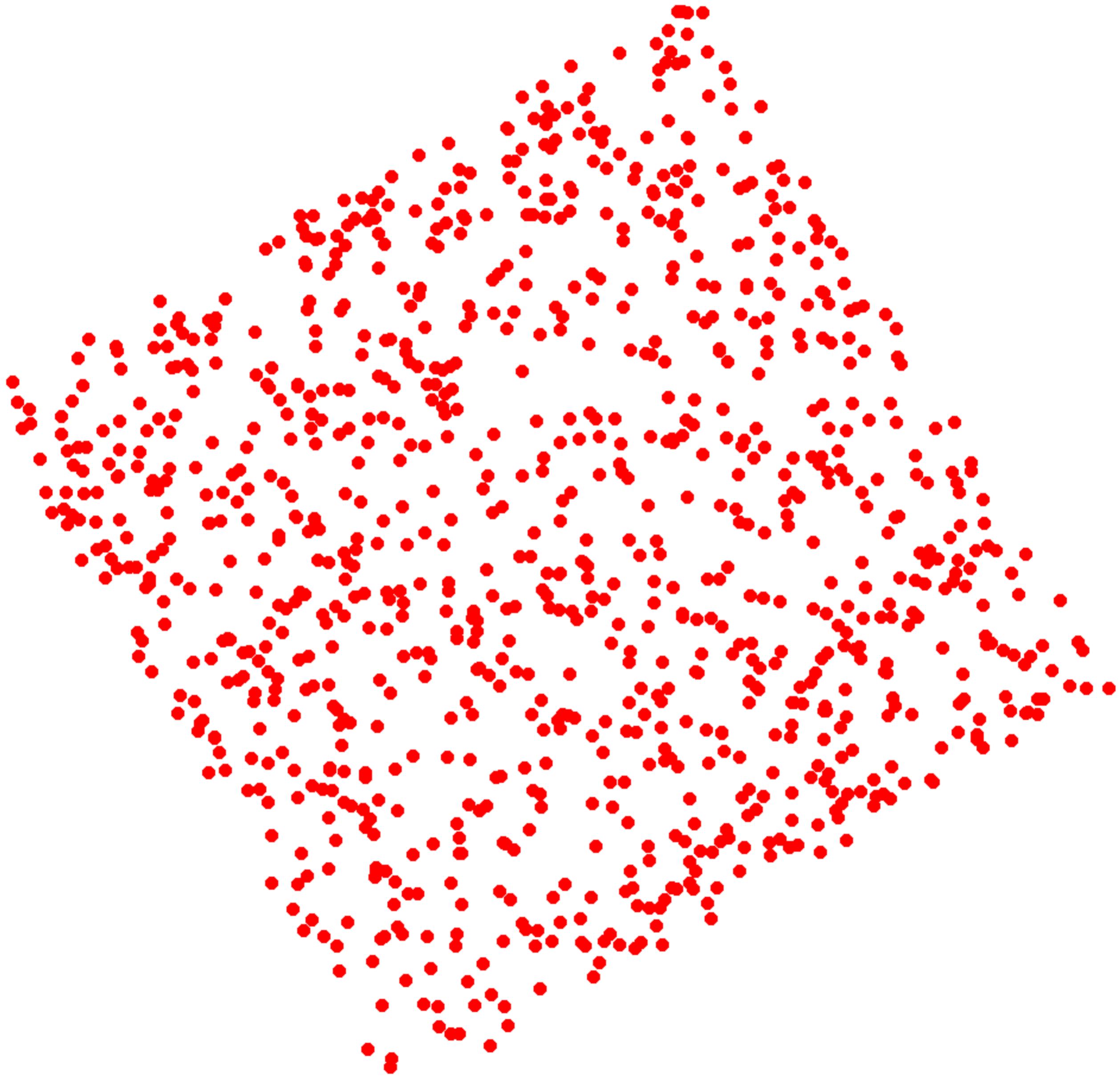
0.00814



-0.01797



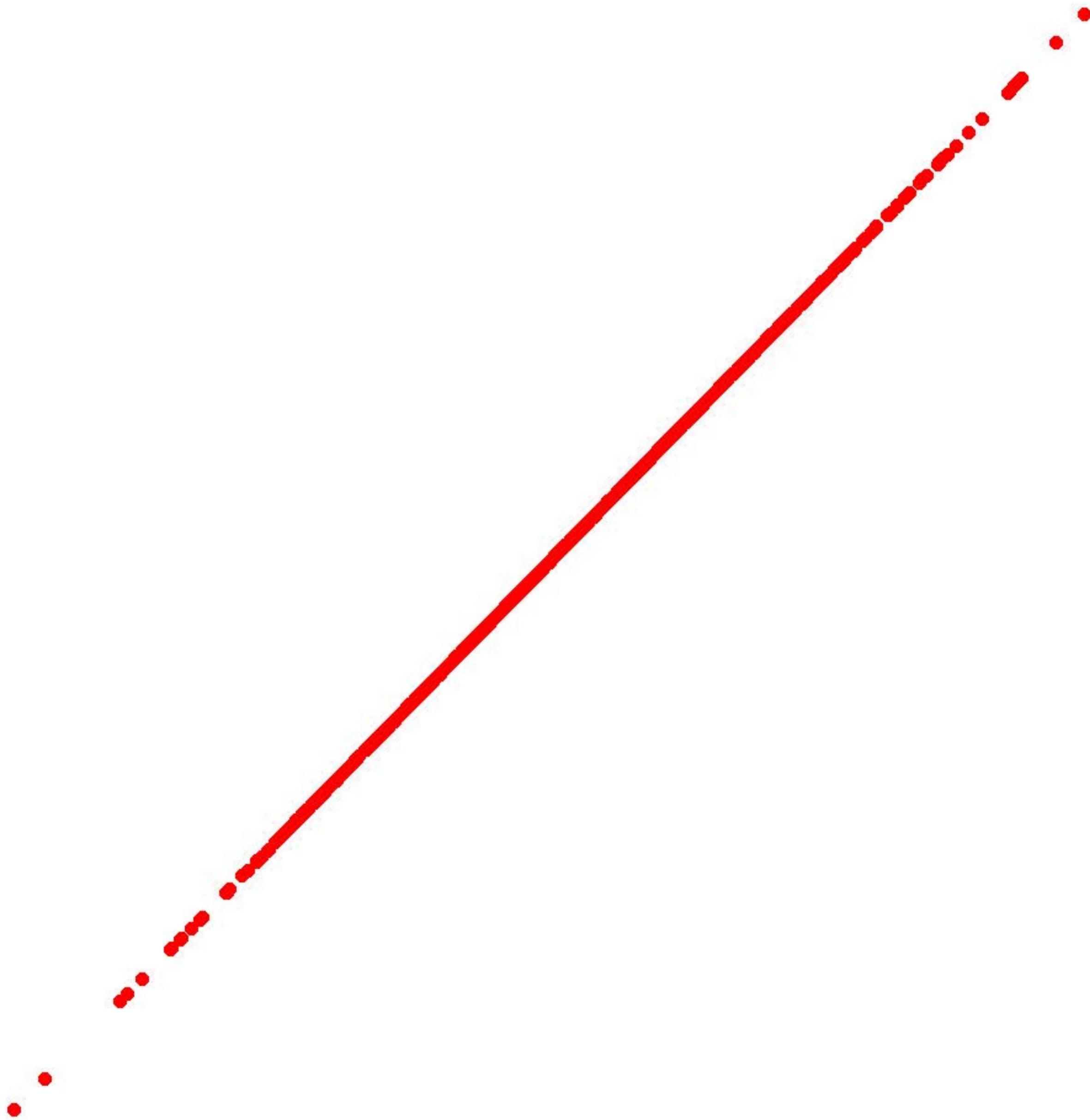
-0.04635



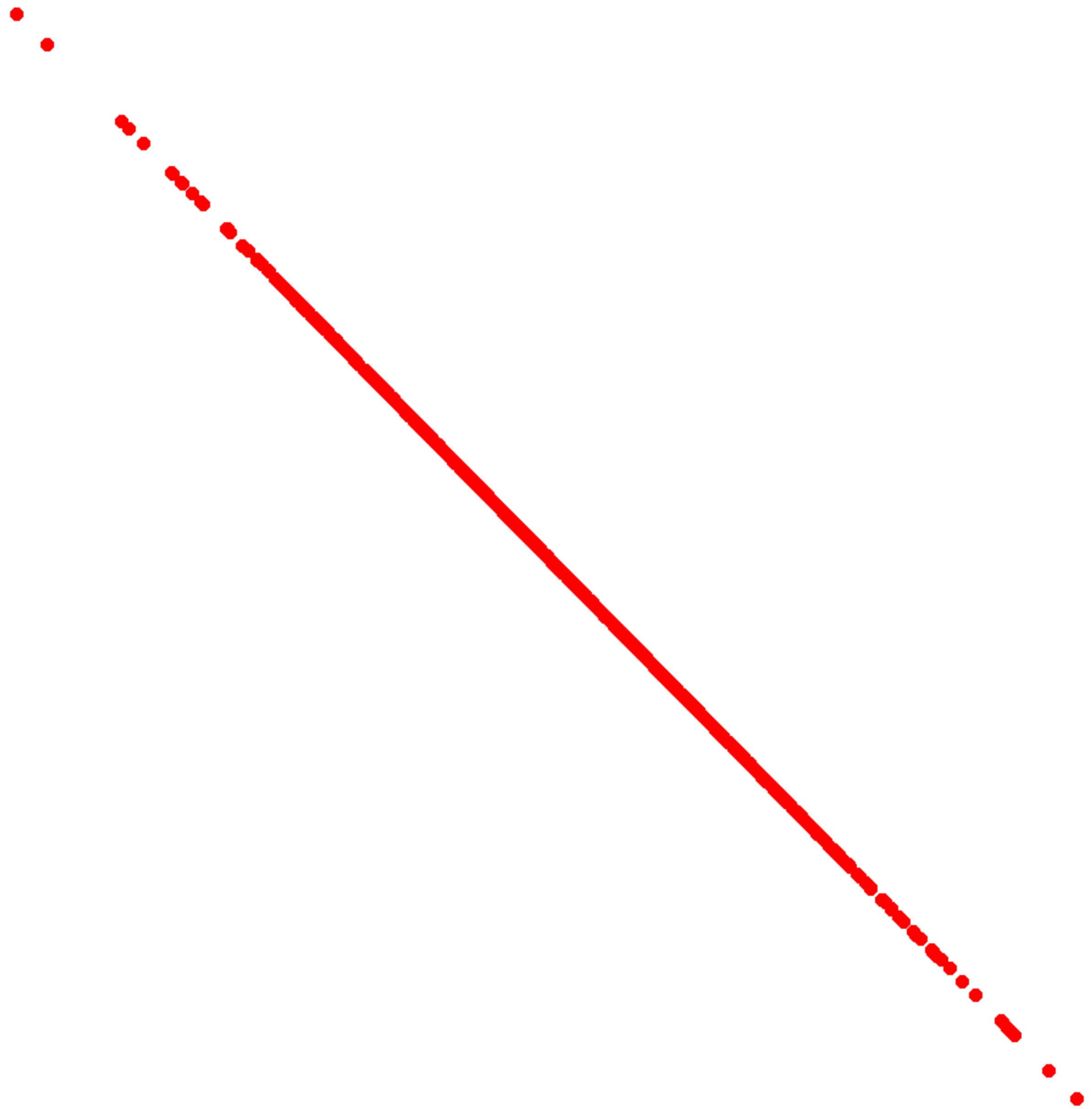
0.01869



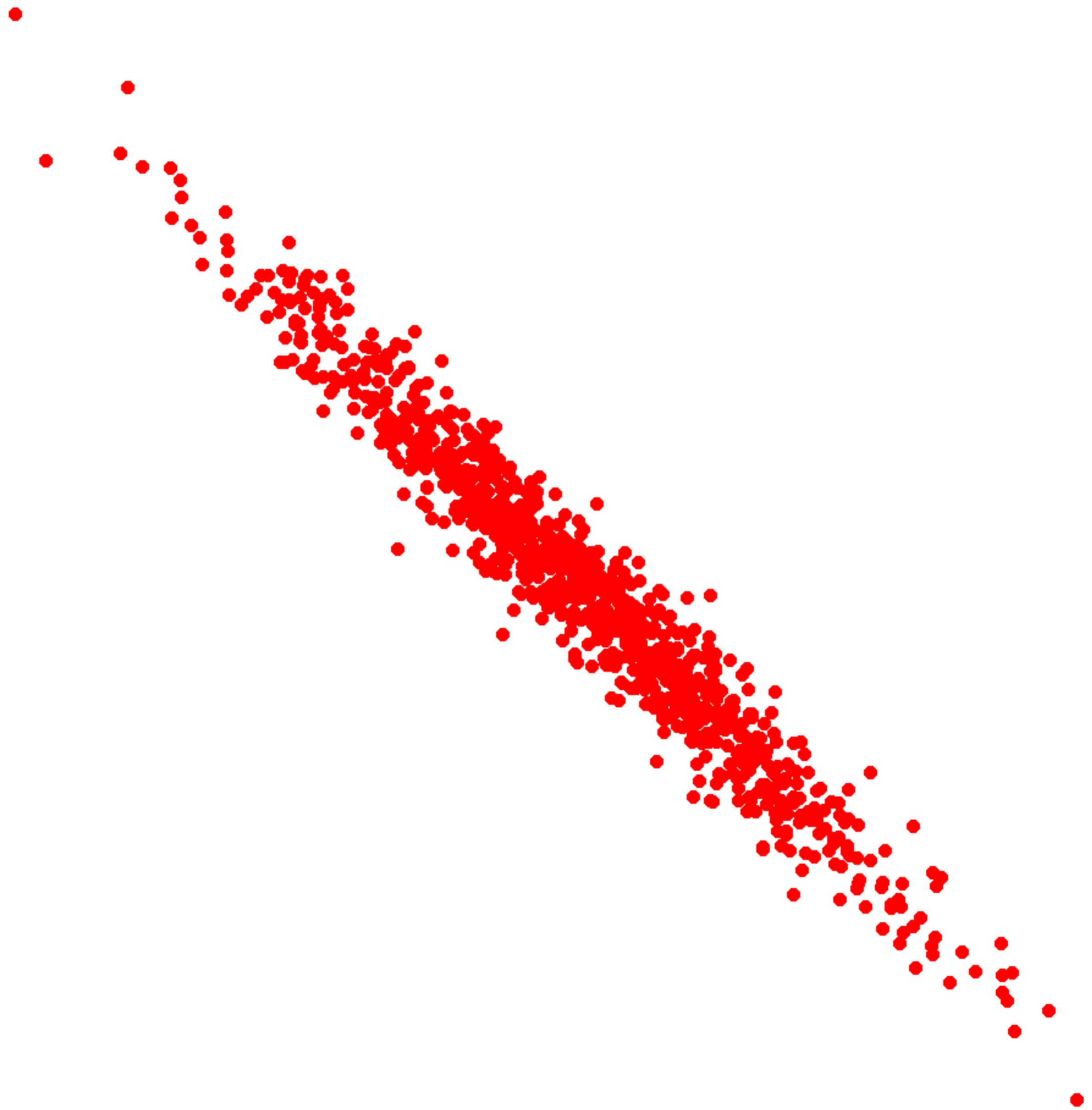
1



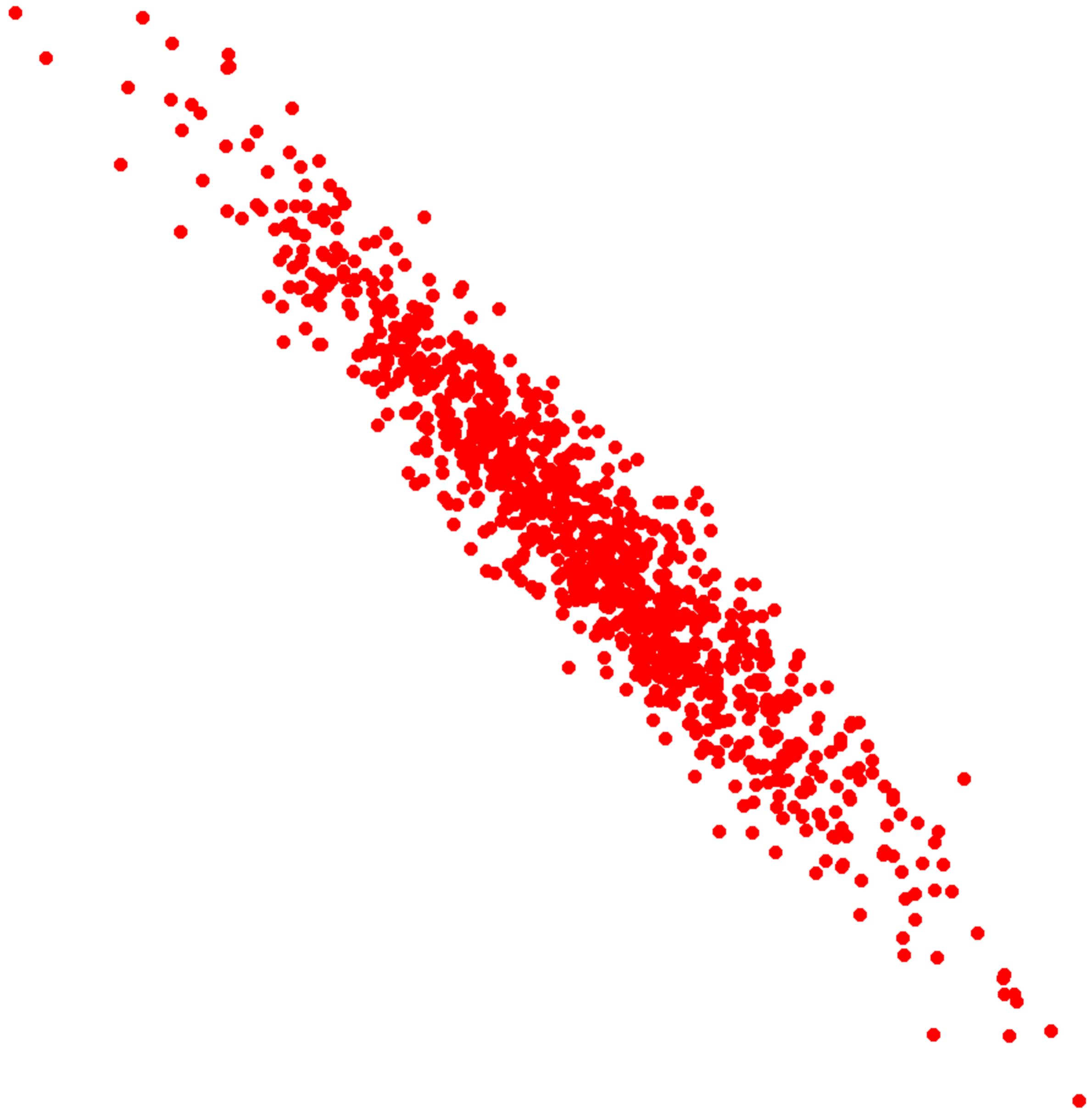
-1



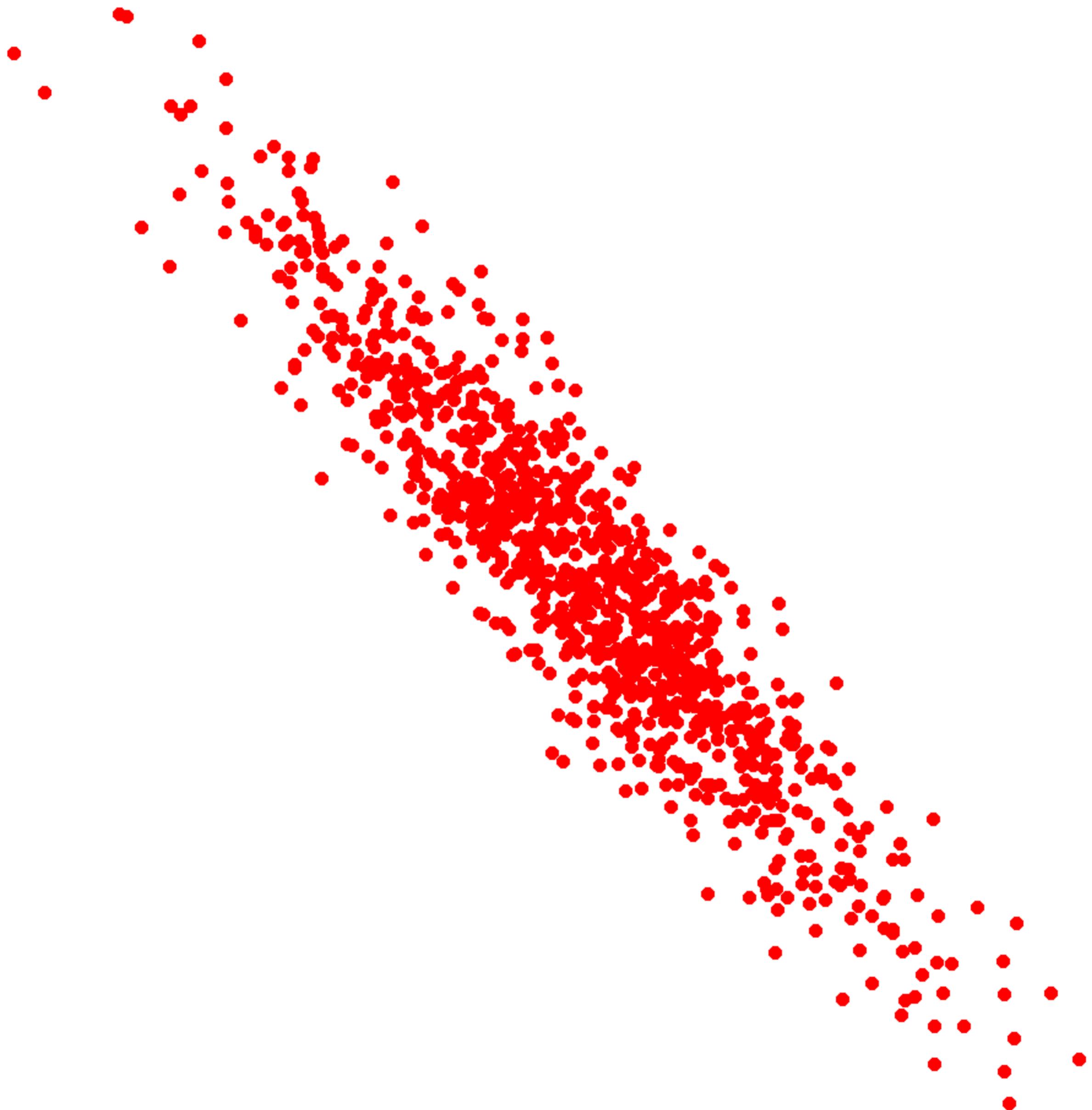
-0.97



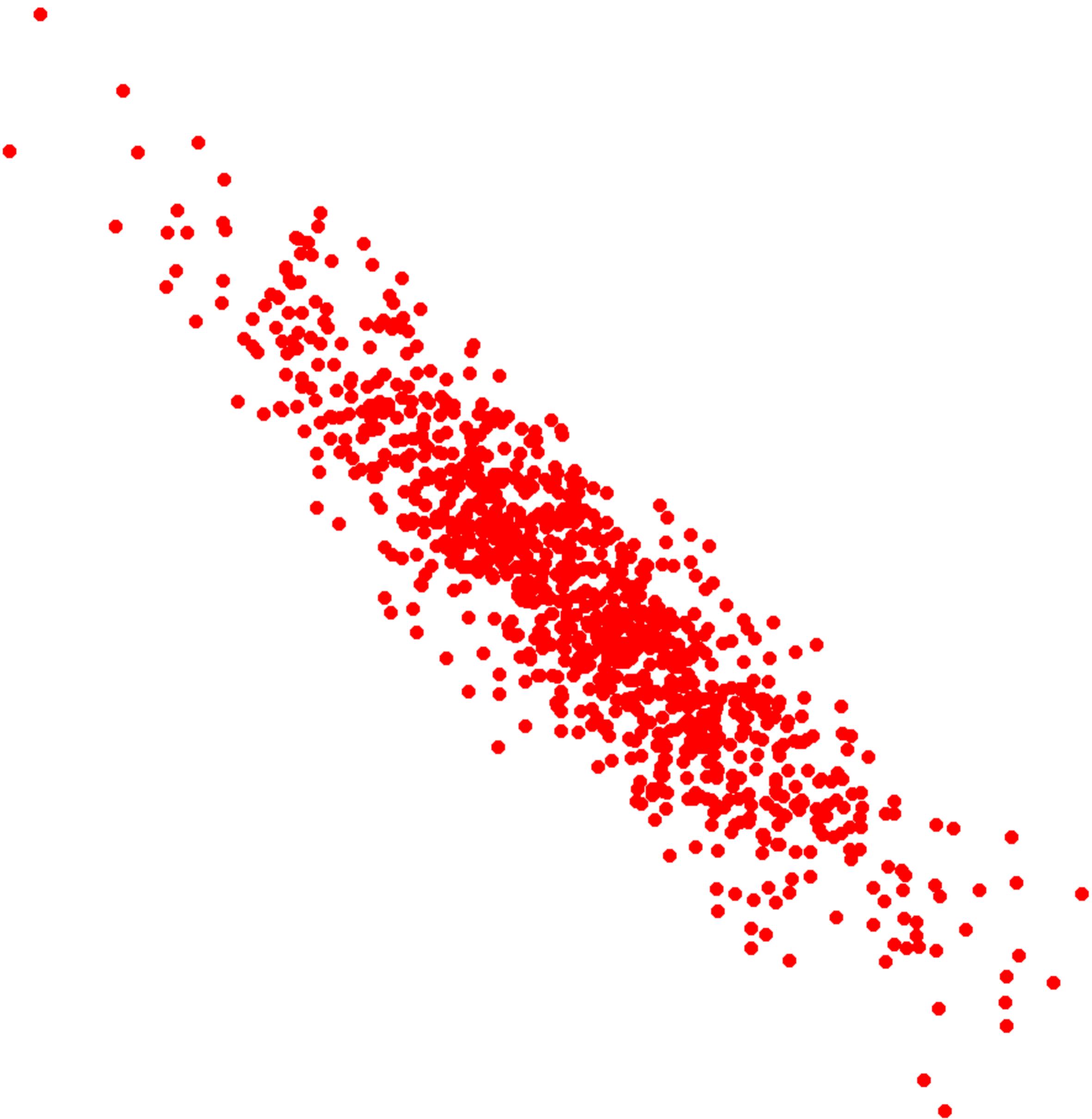
-0.94



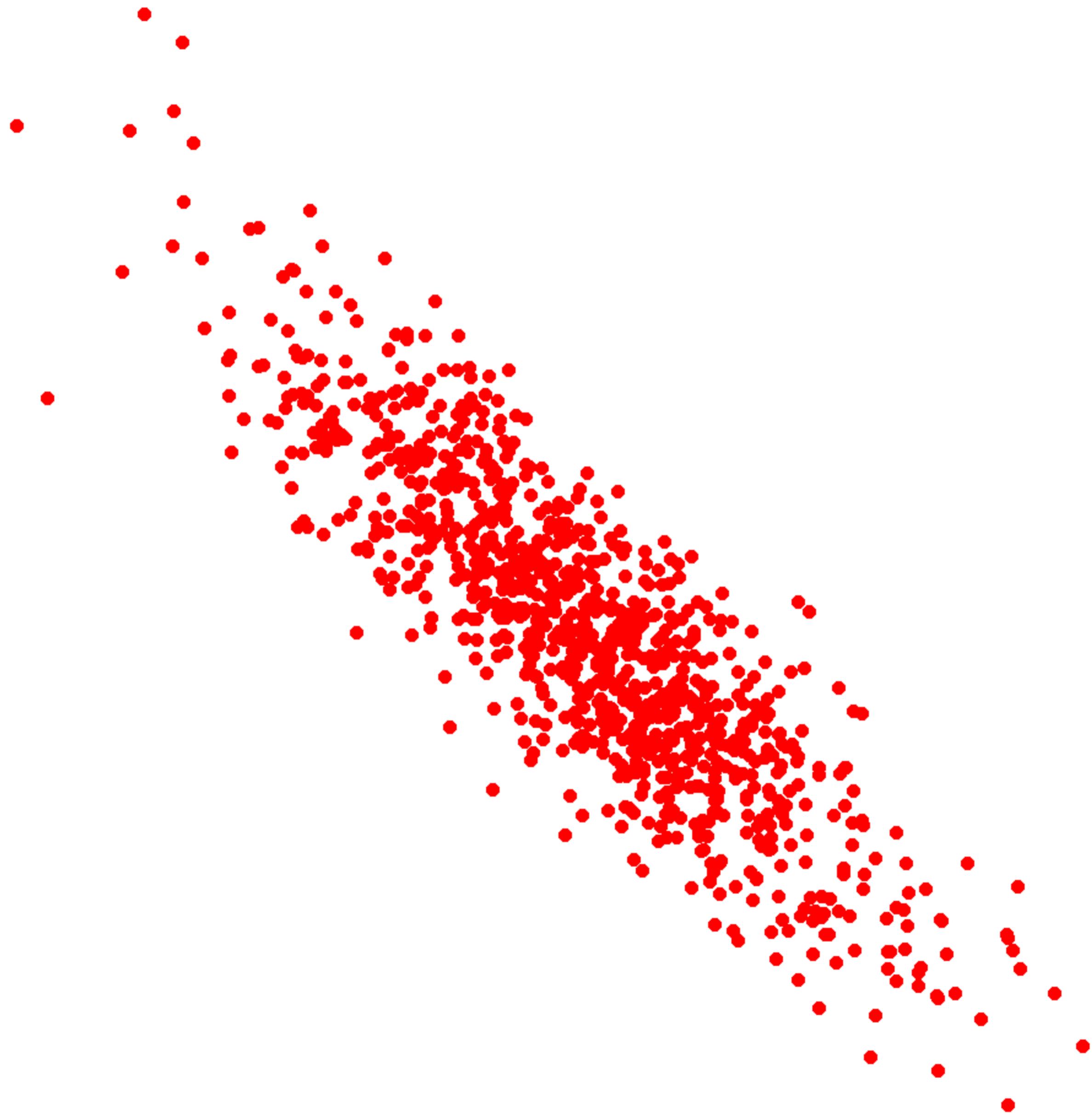
-0.91



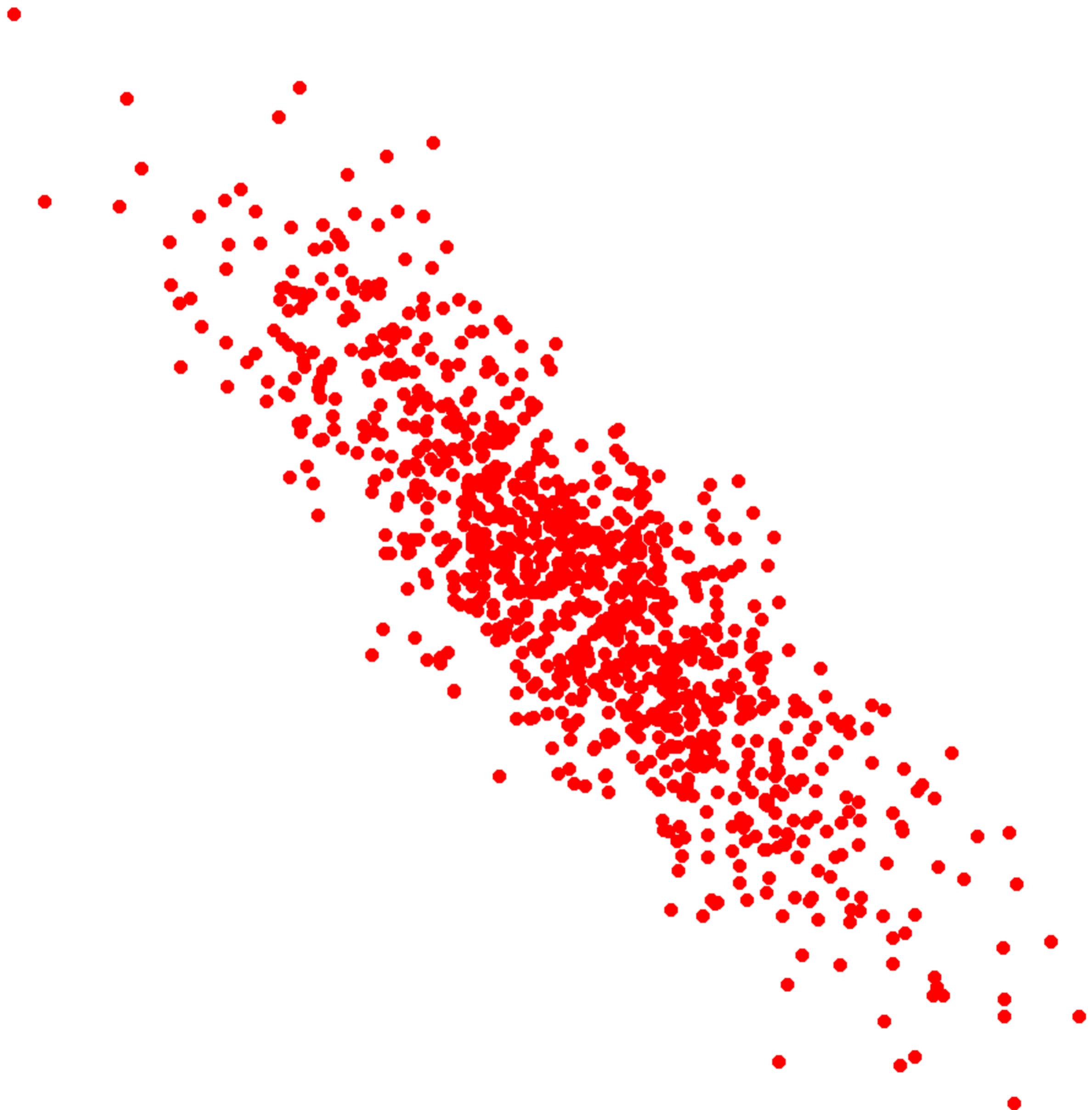
-0.89



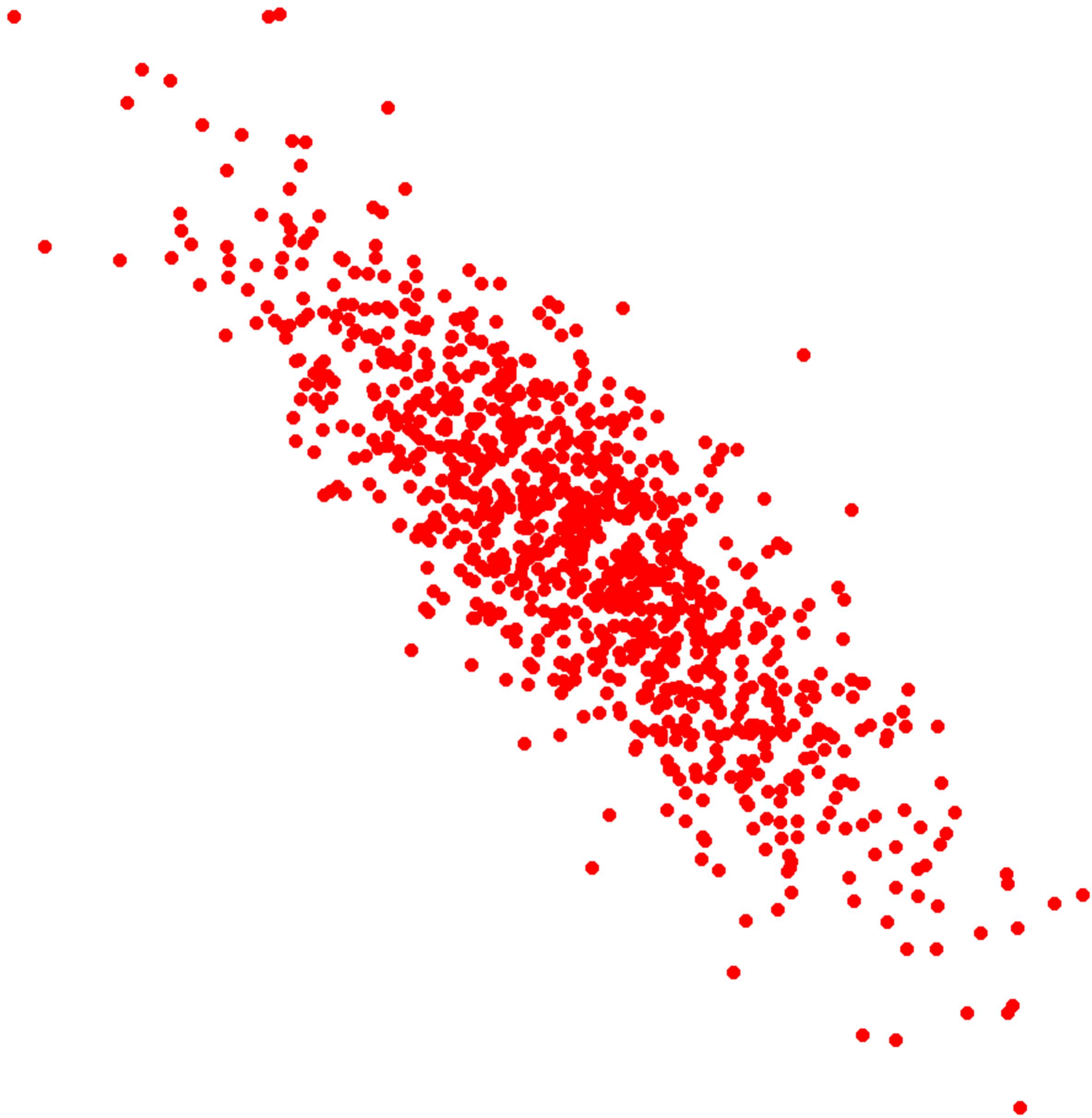
-0.87



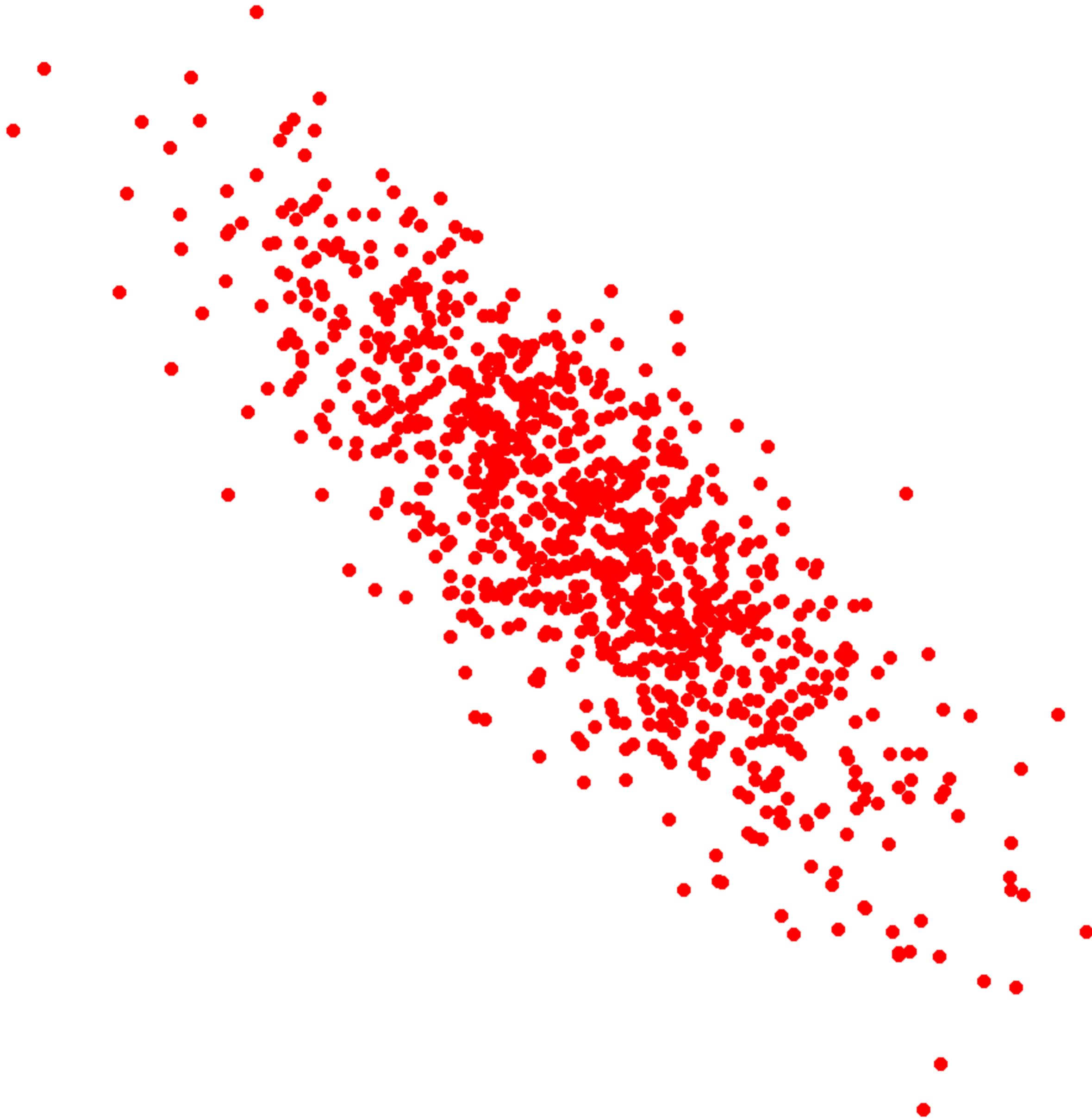
-0.82



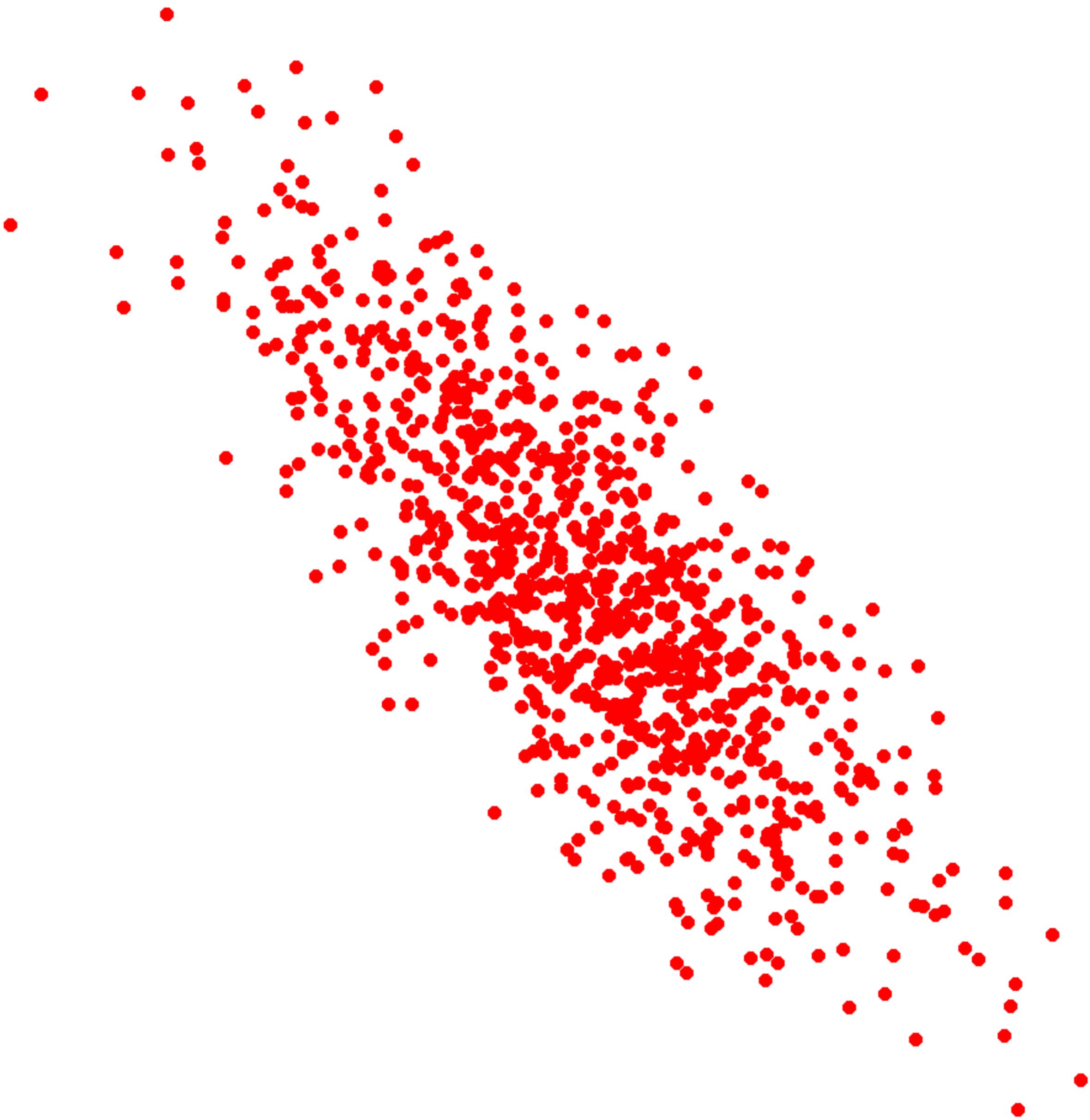
-0.81



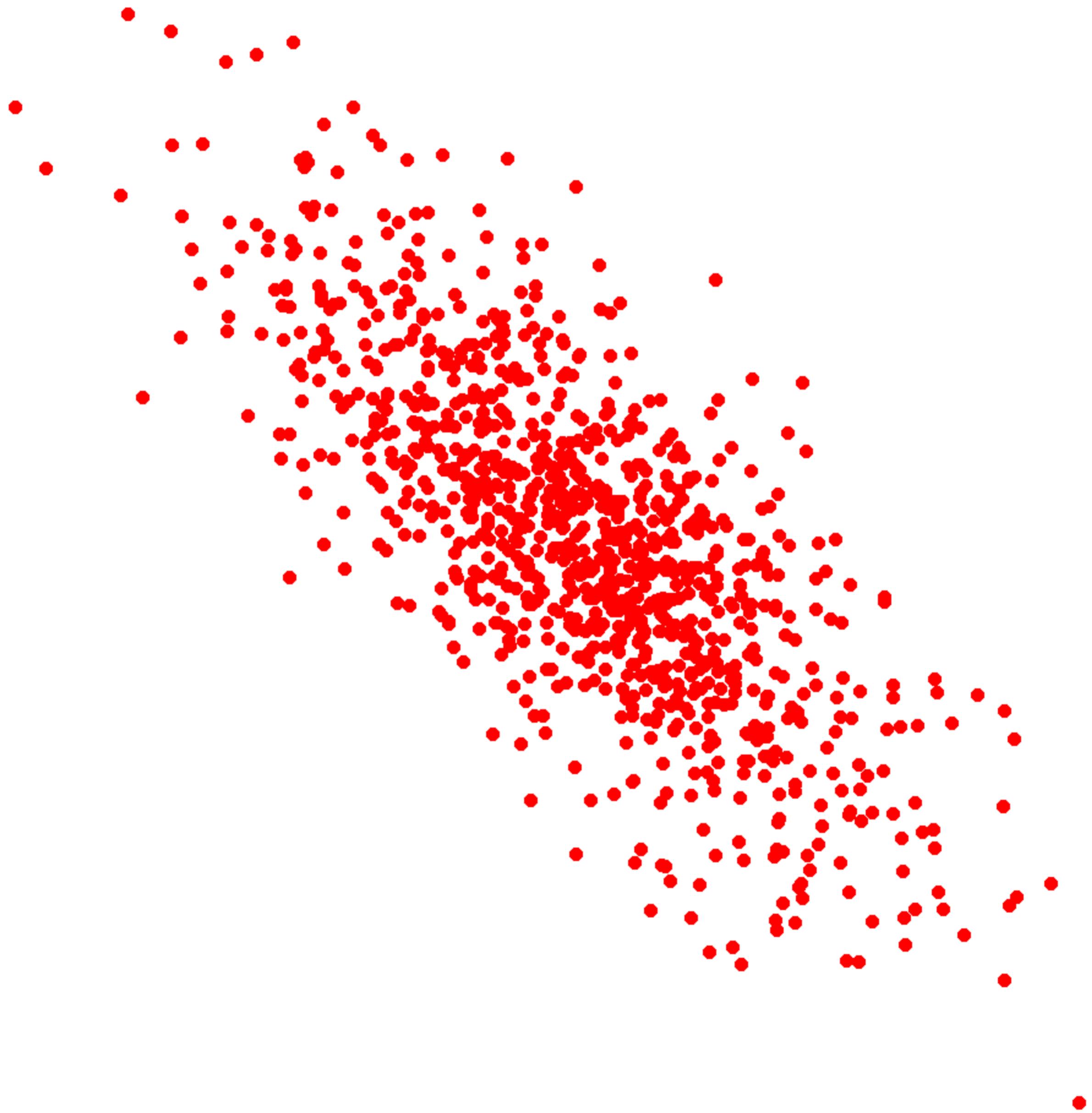
-0.79



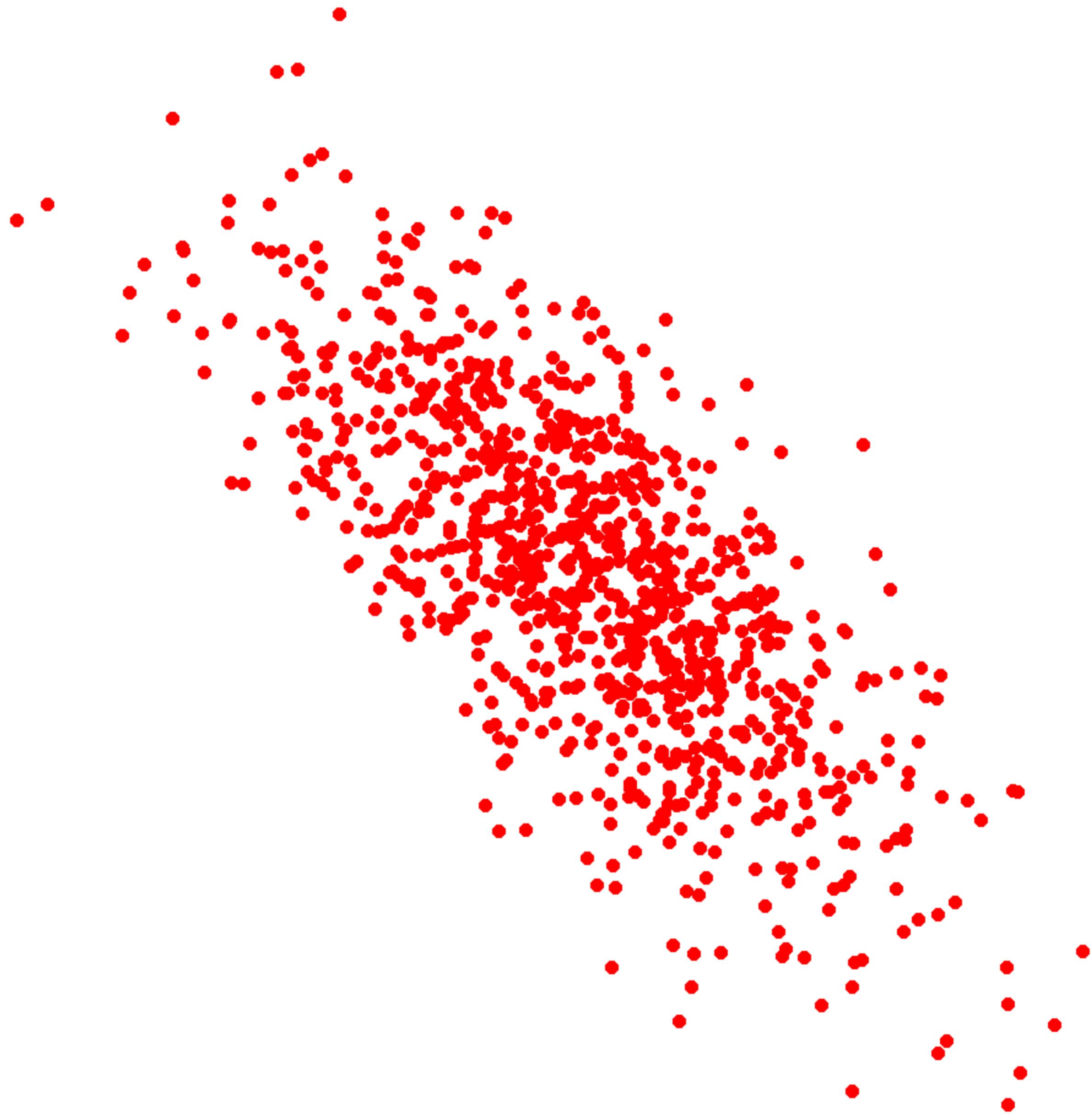
-0.76



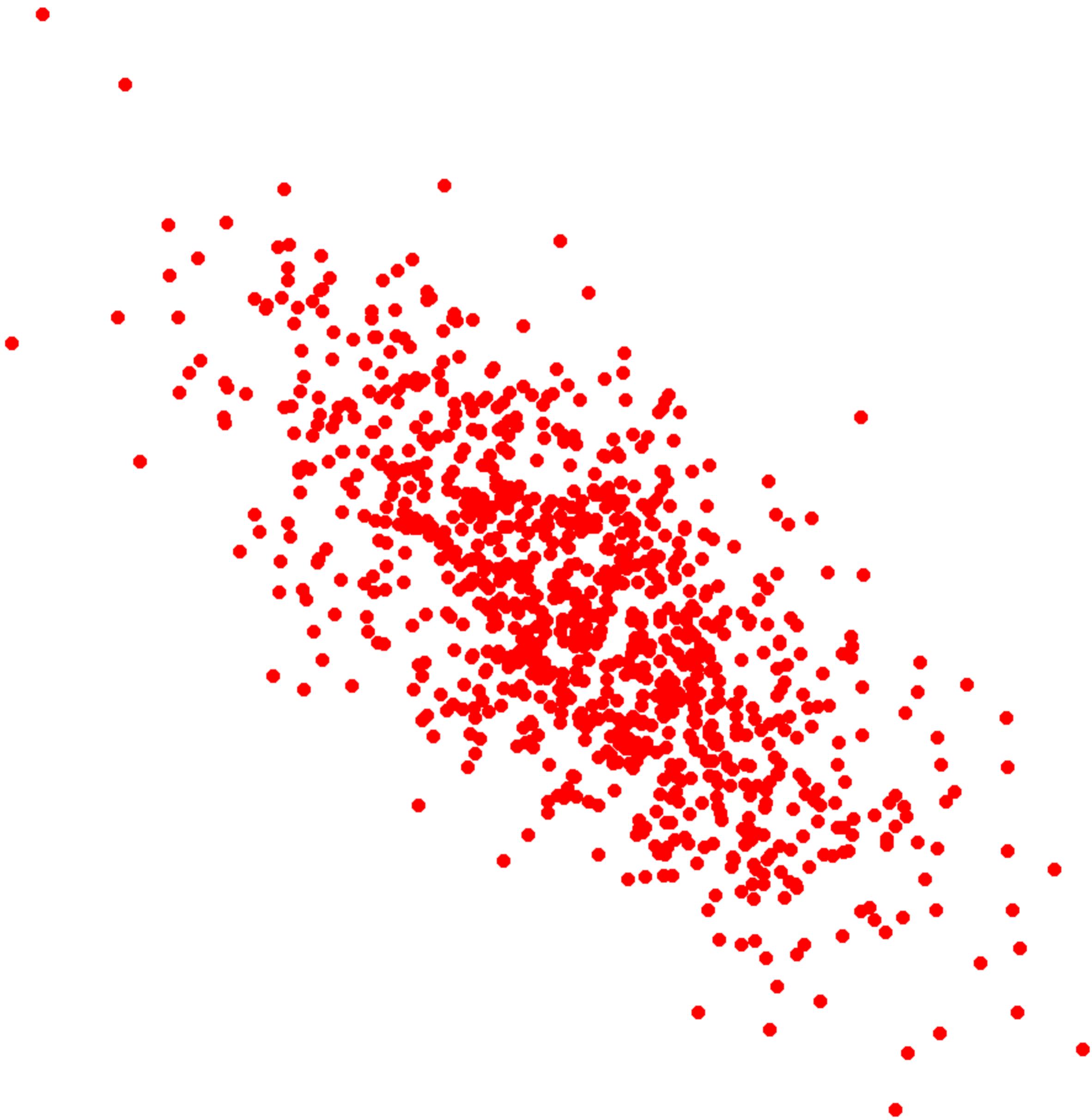
-0.74



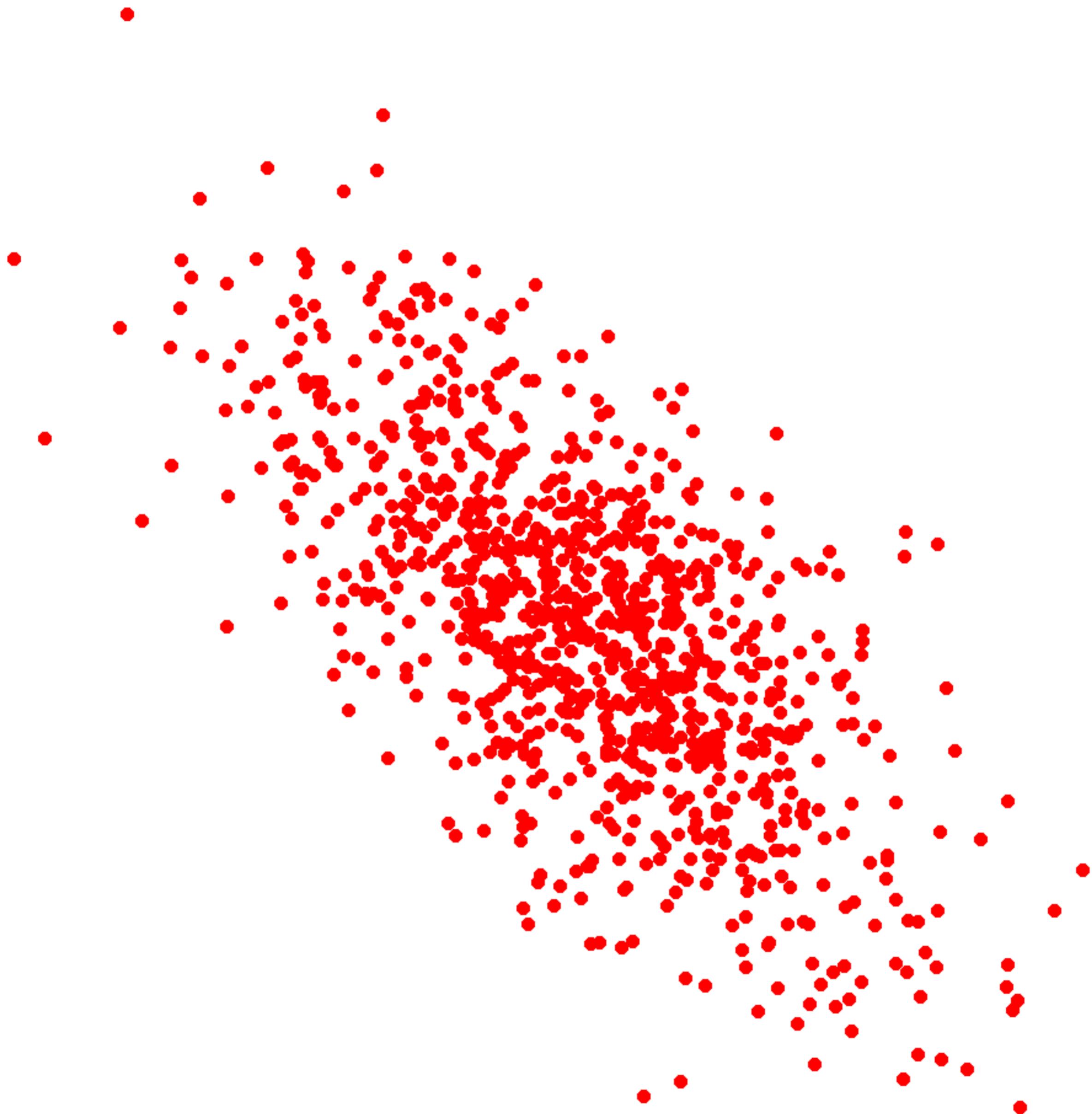
-0.71



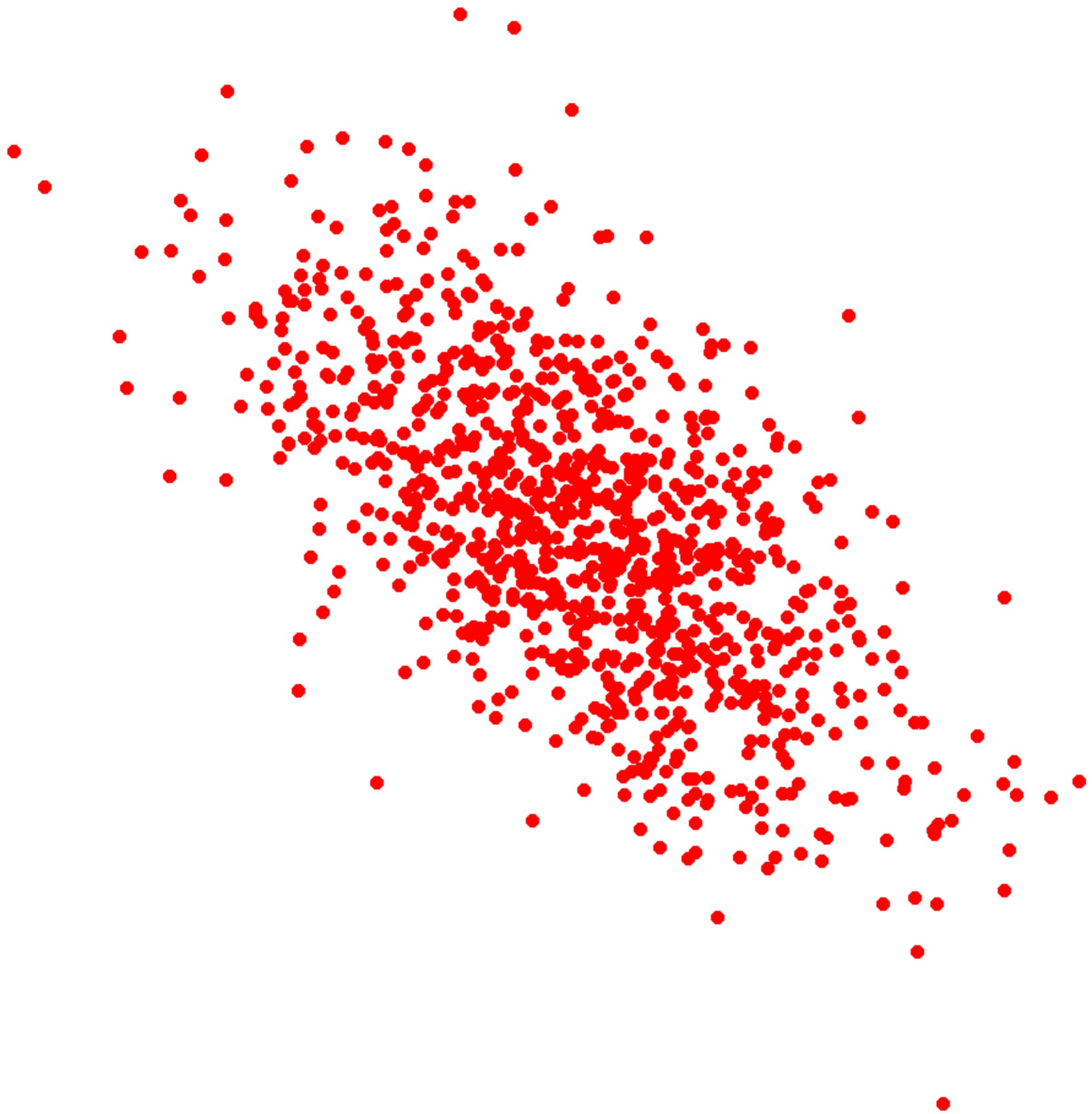
-0.68



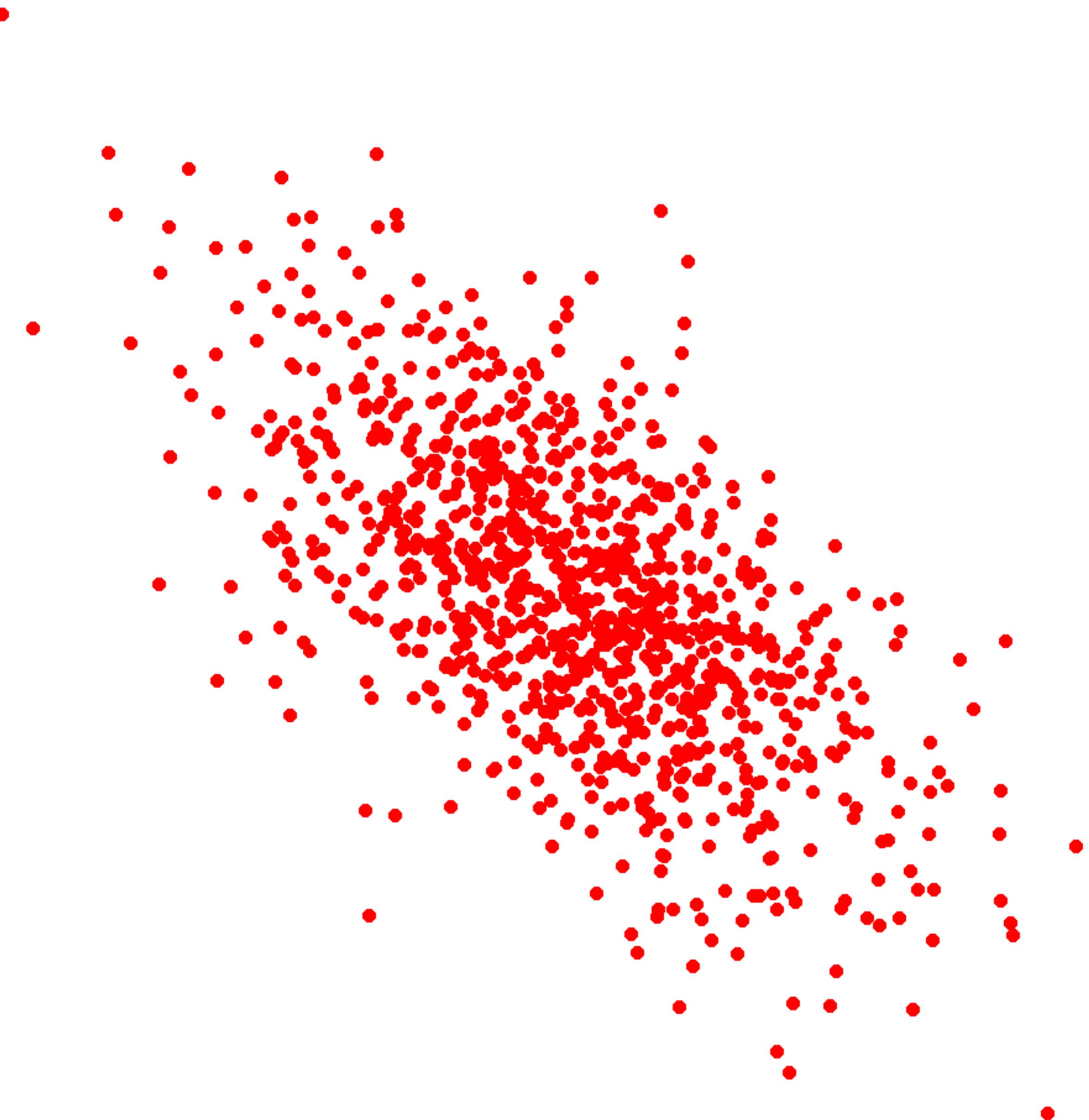
-0.65



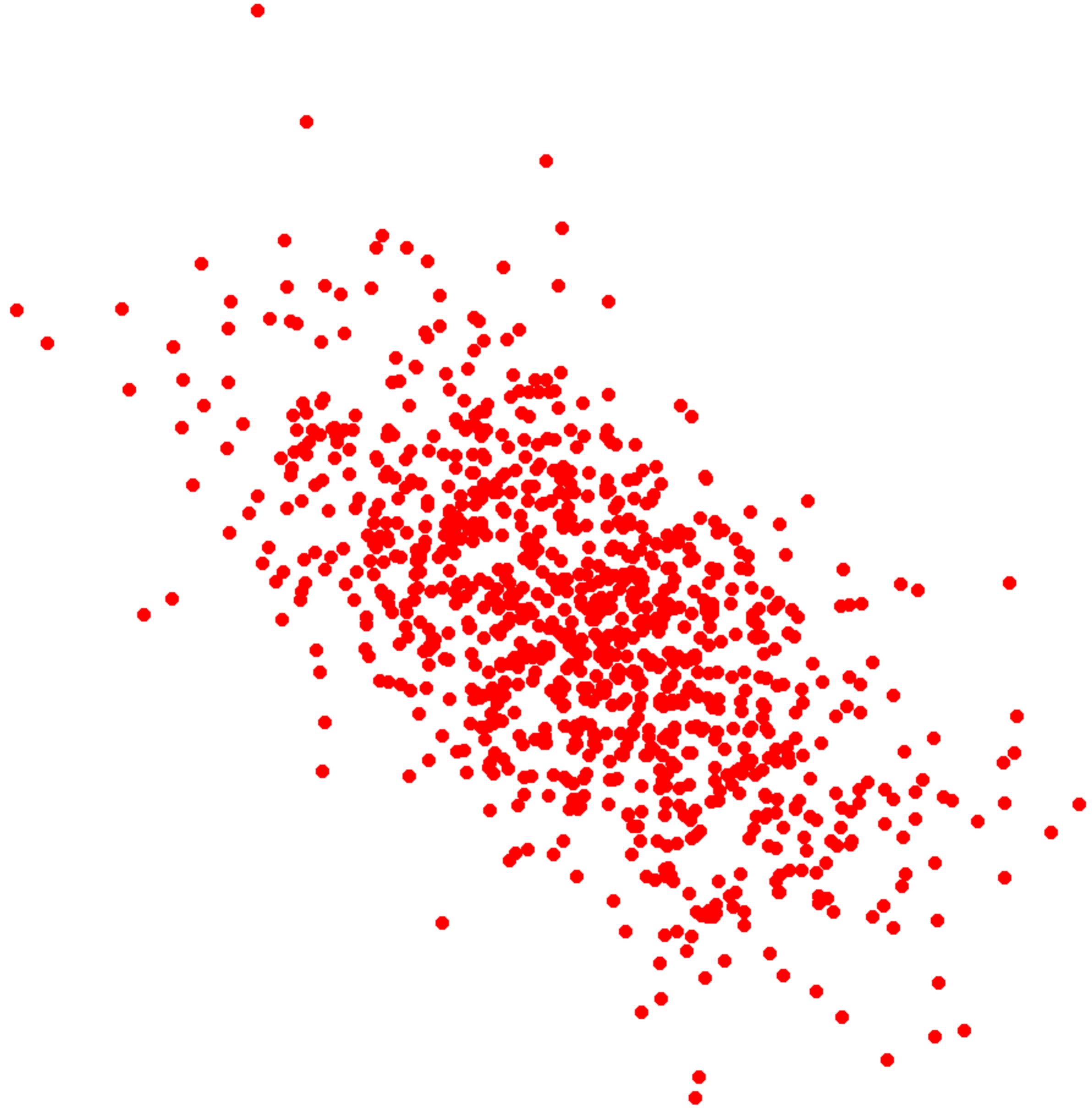
-0.61



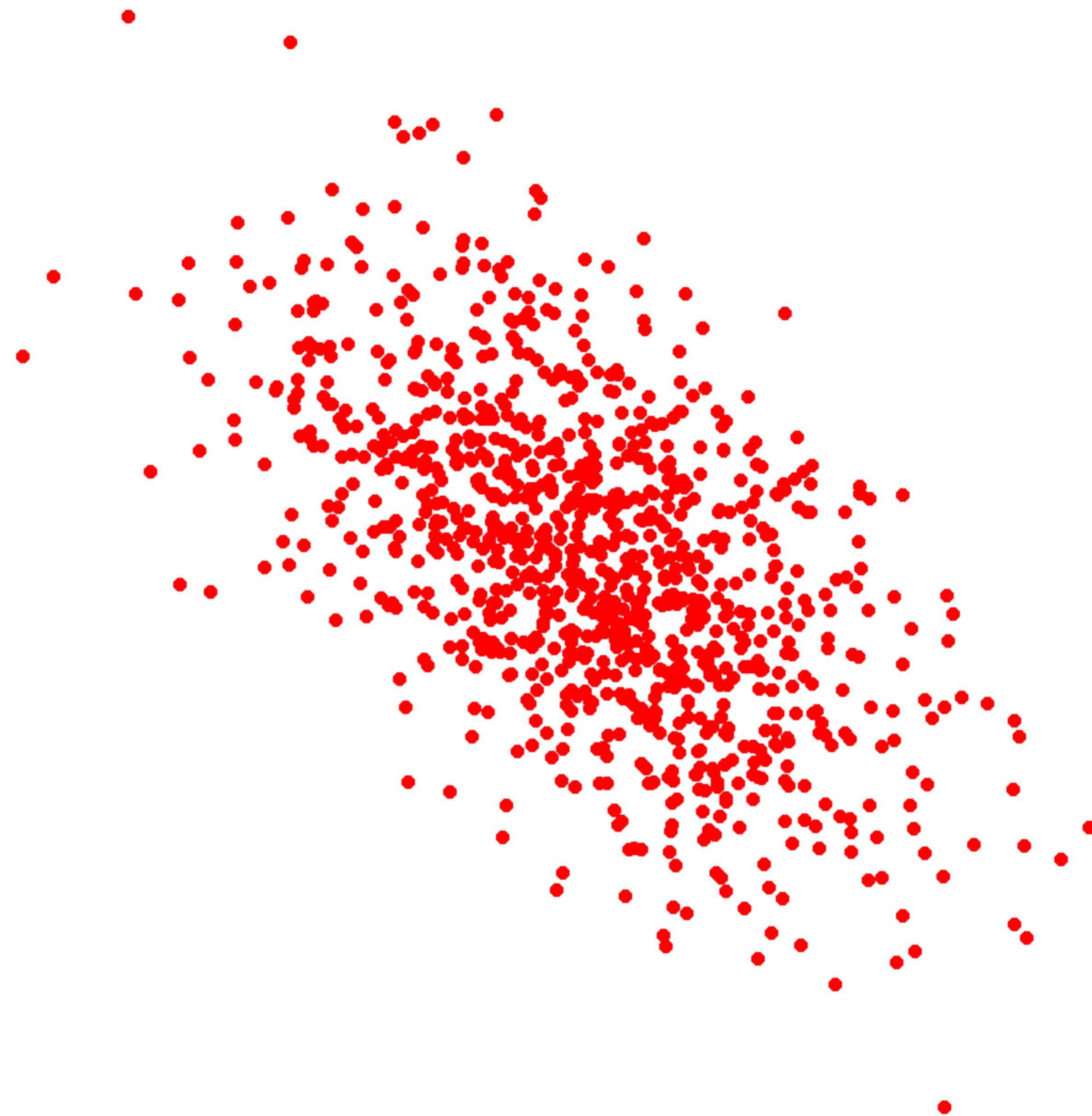
-0.6



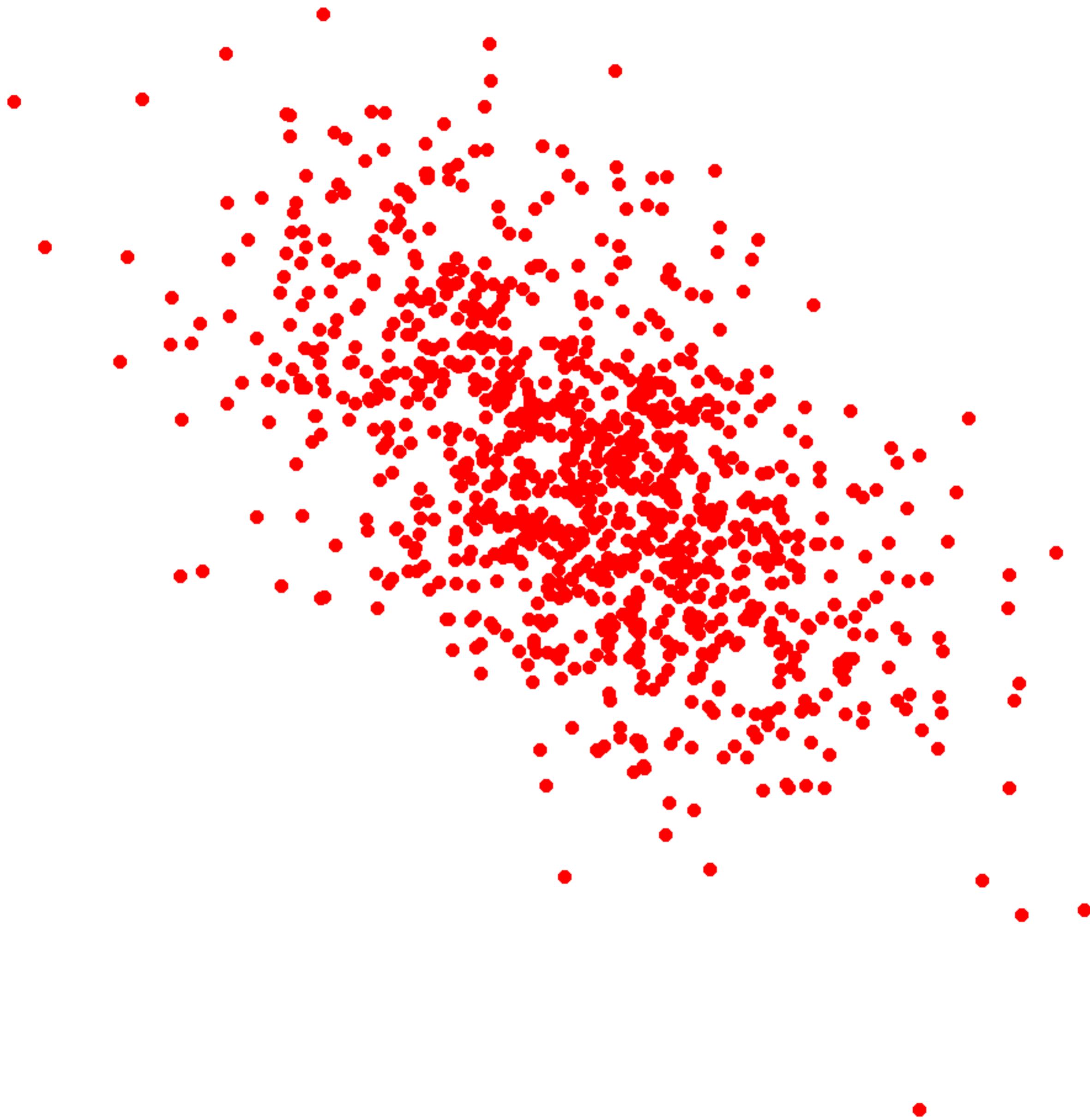
-0.59



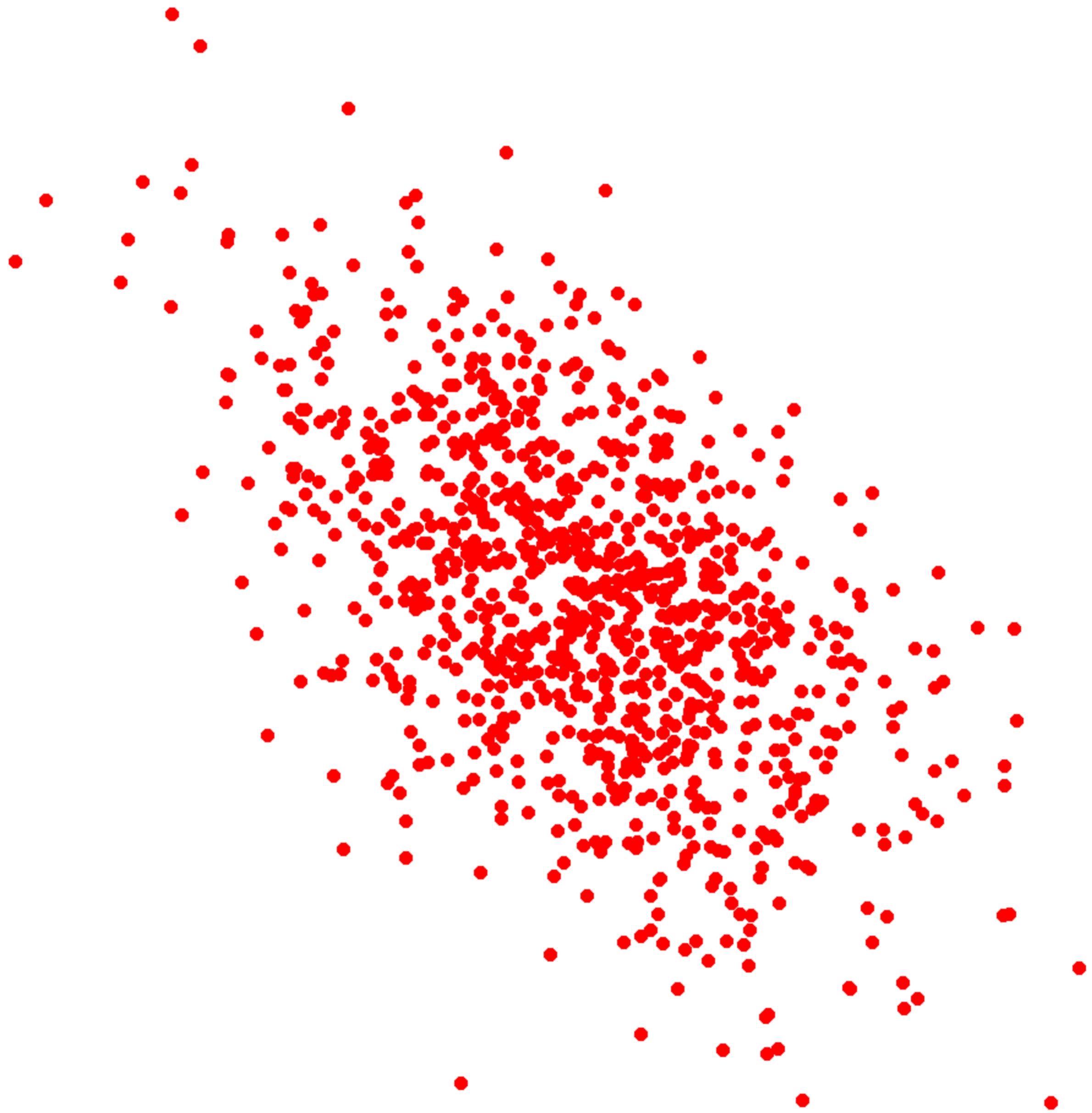
-0.57



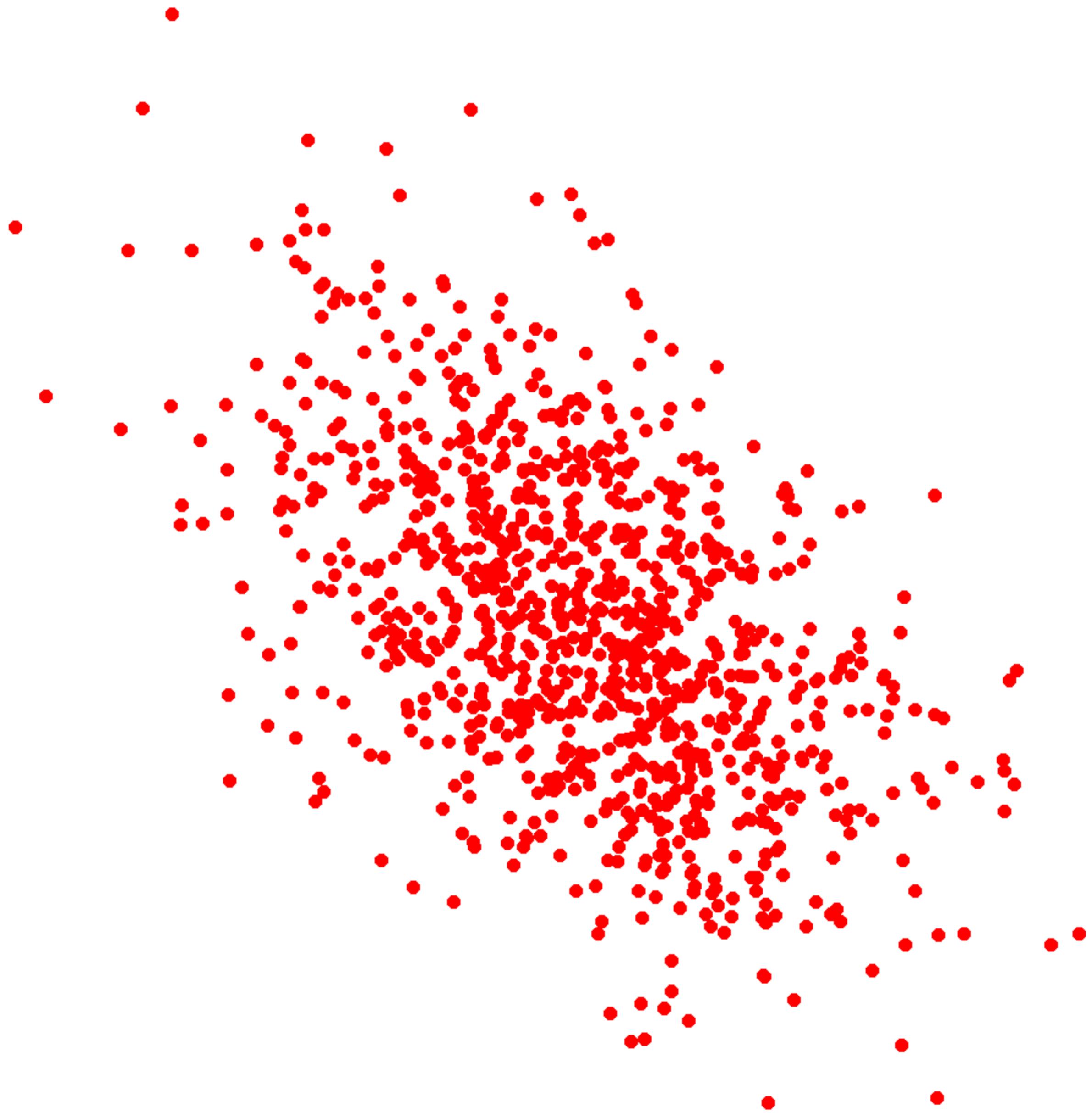
-0.53



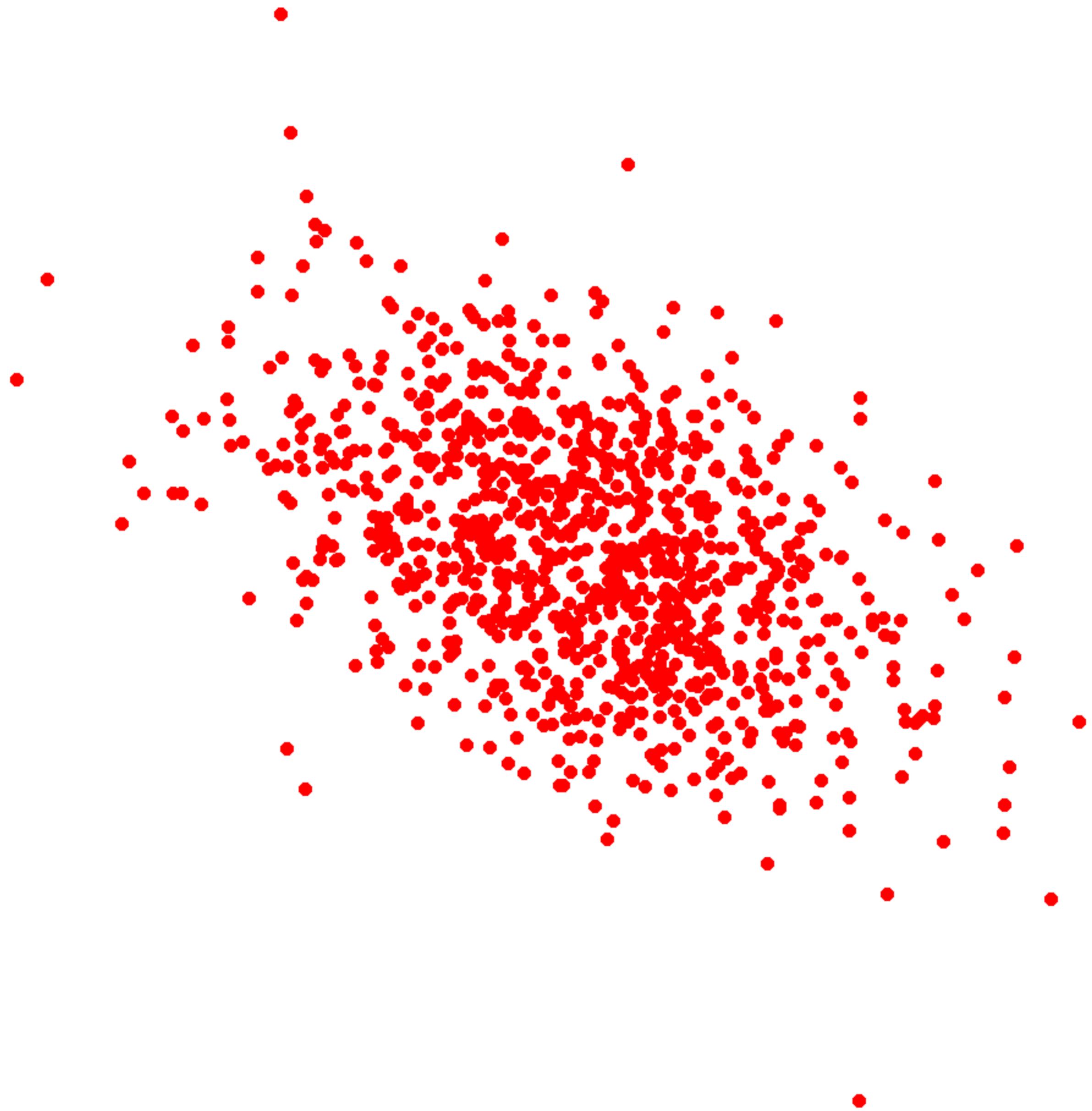
-0.51



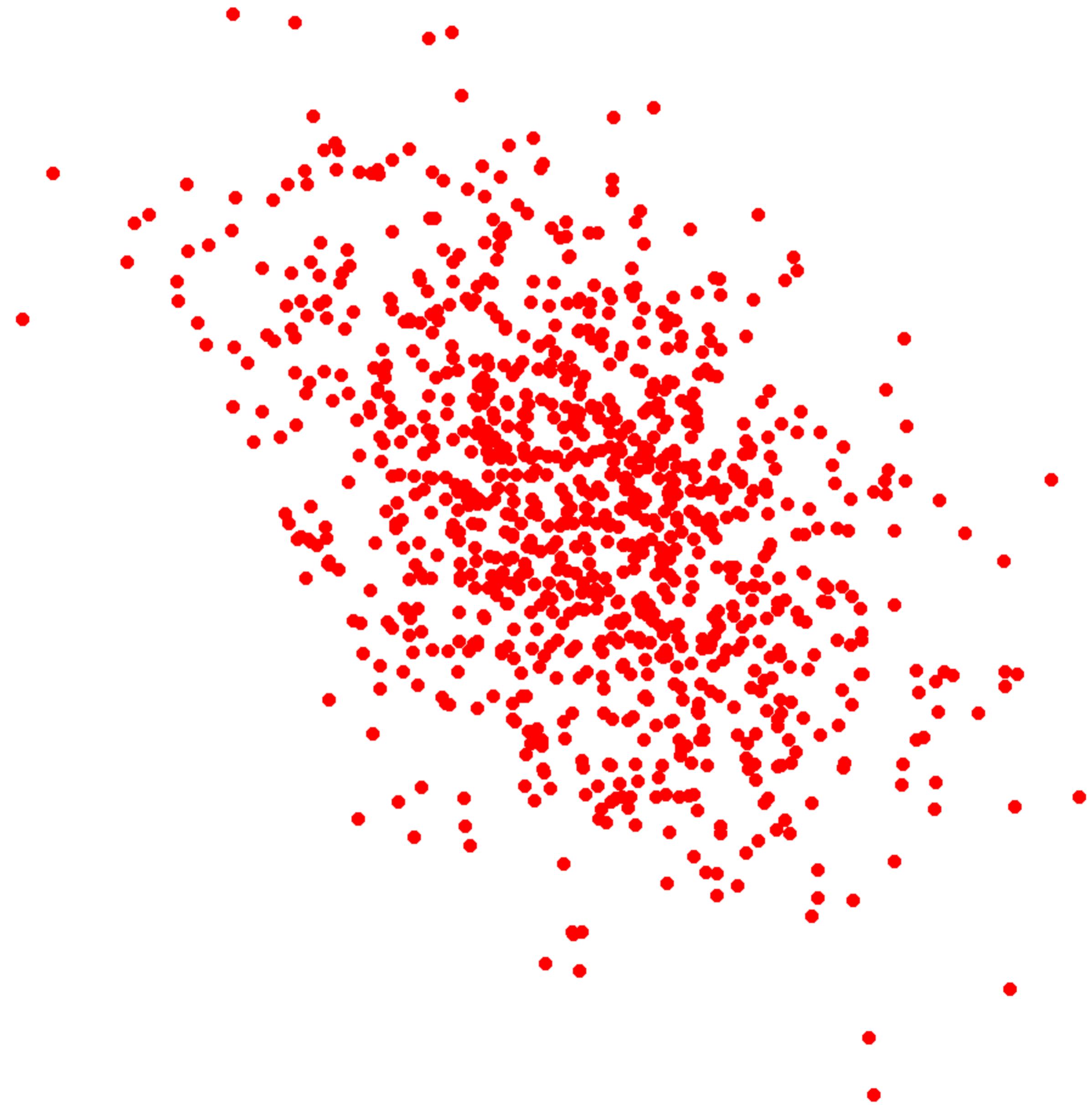
-0.5



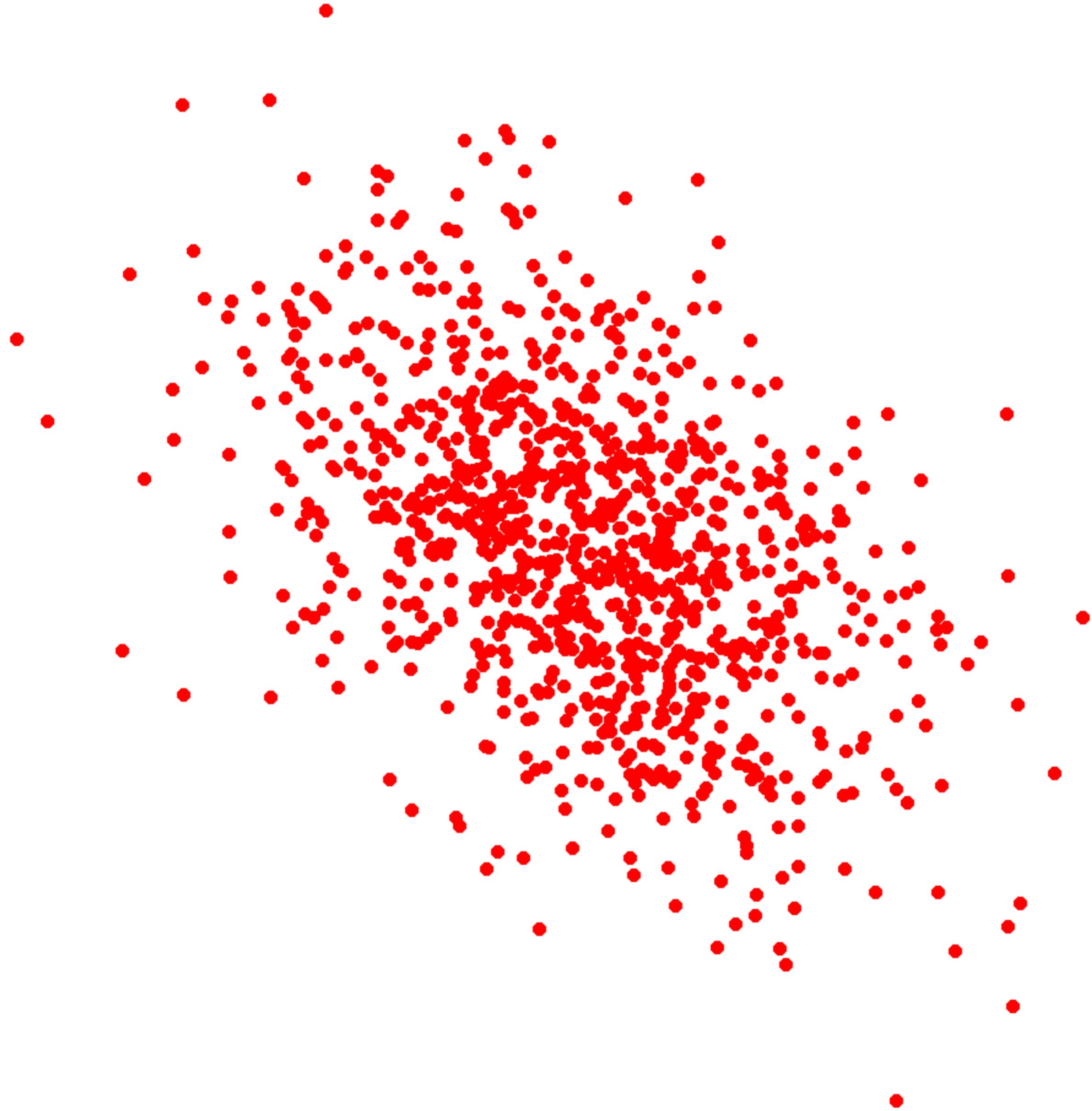
-0.42



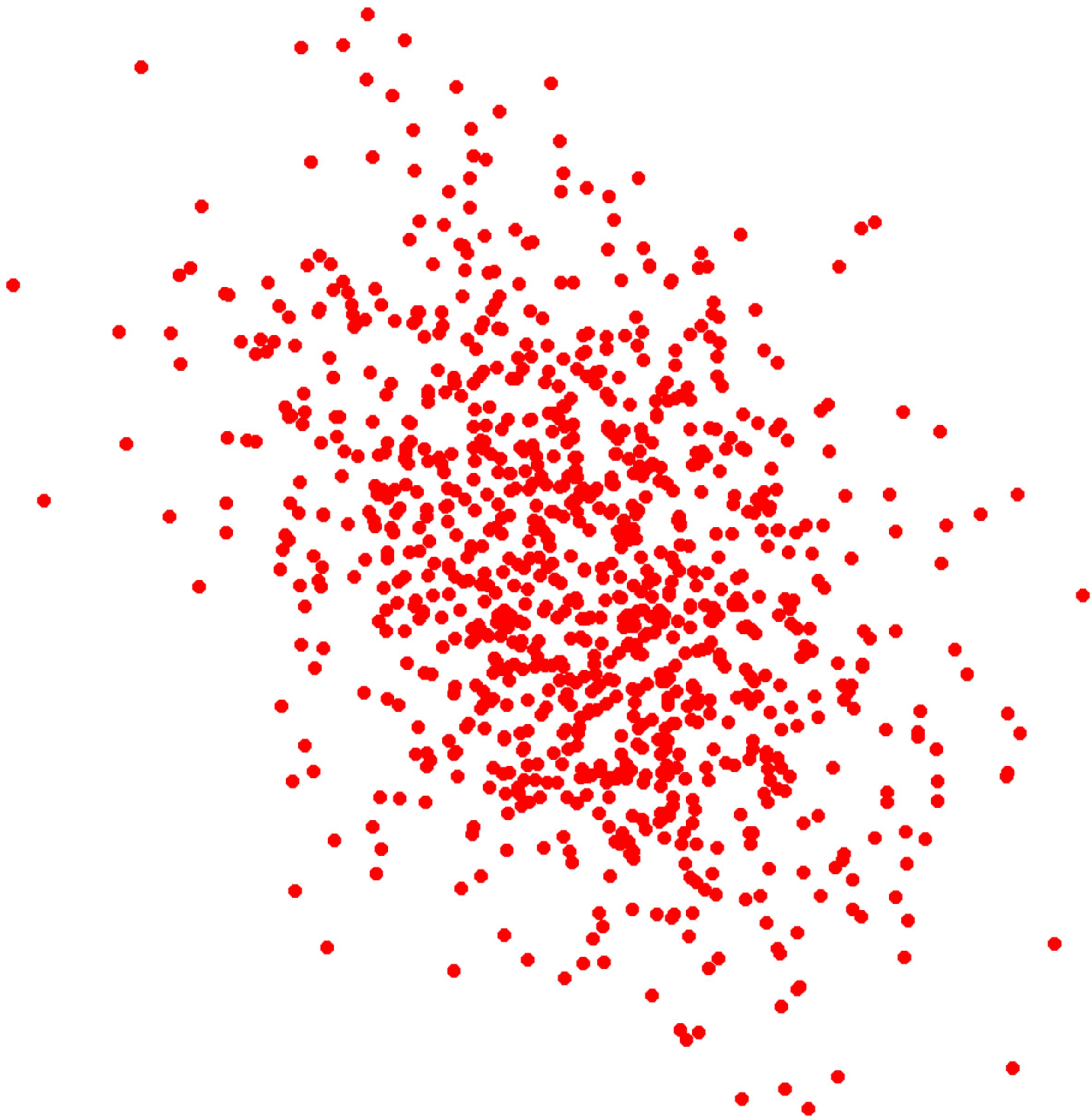
-0.39



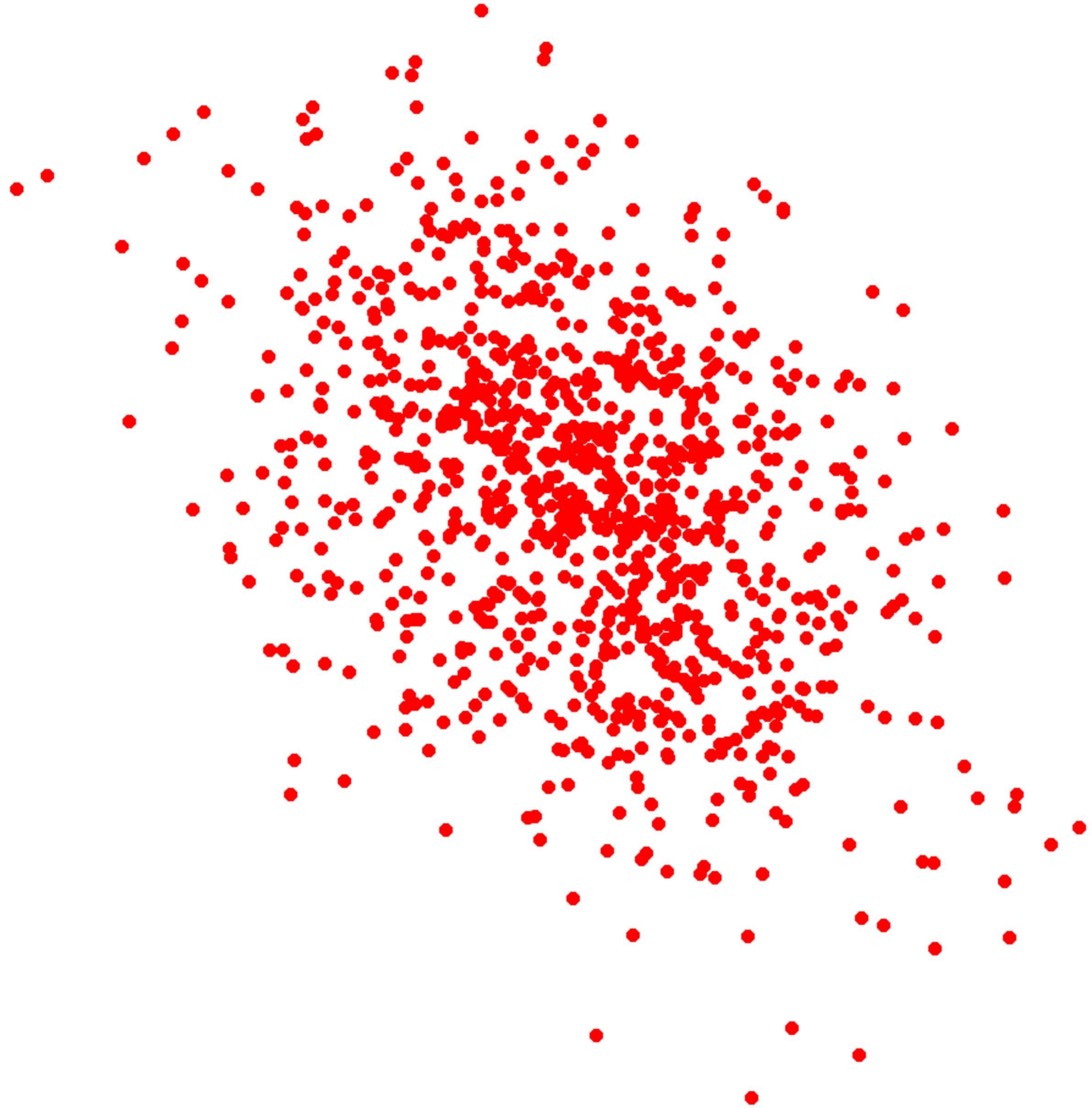
-0.45



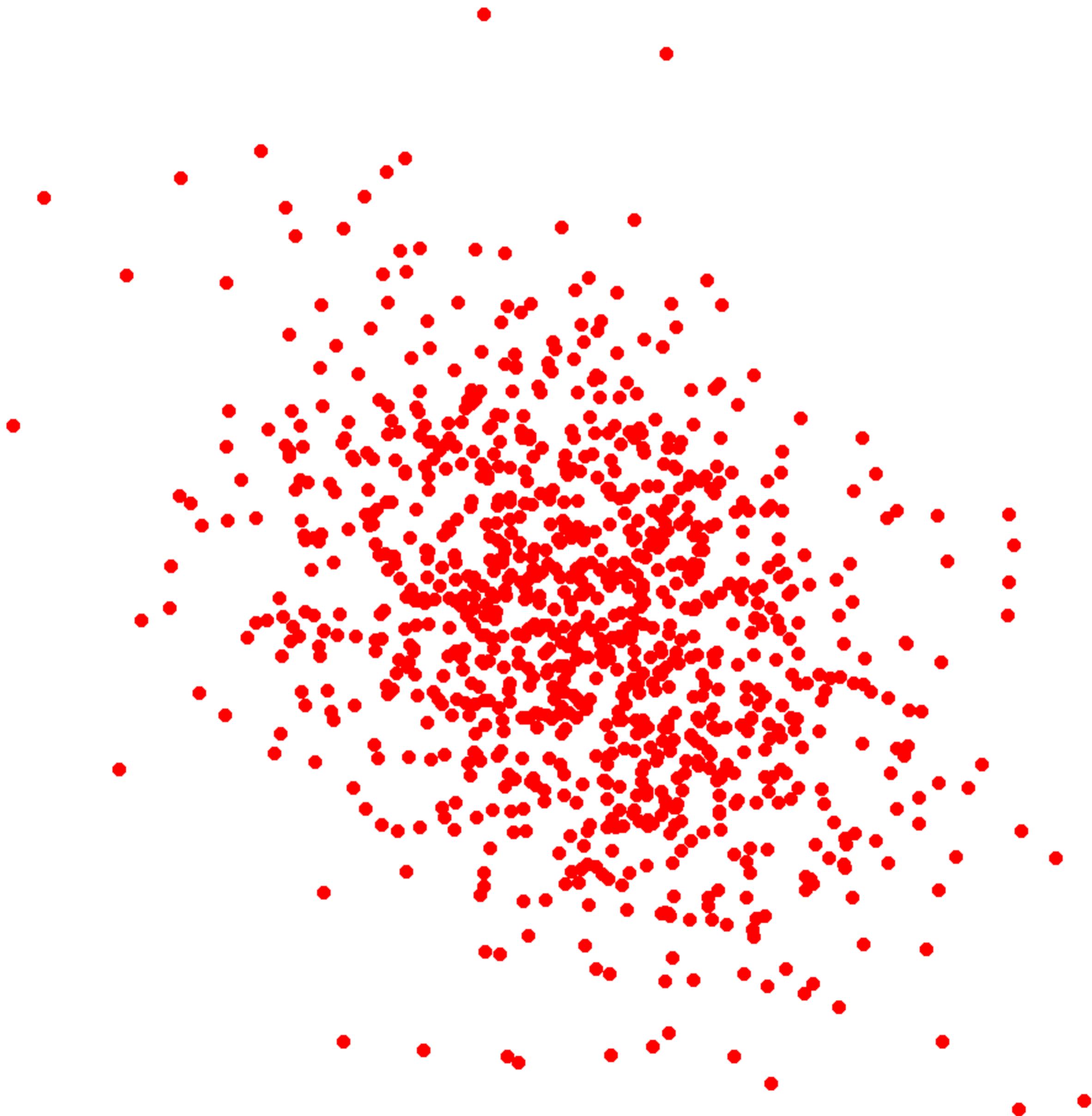
-0.35



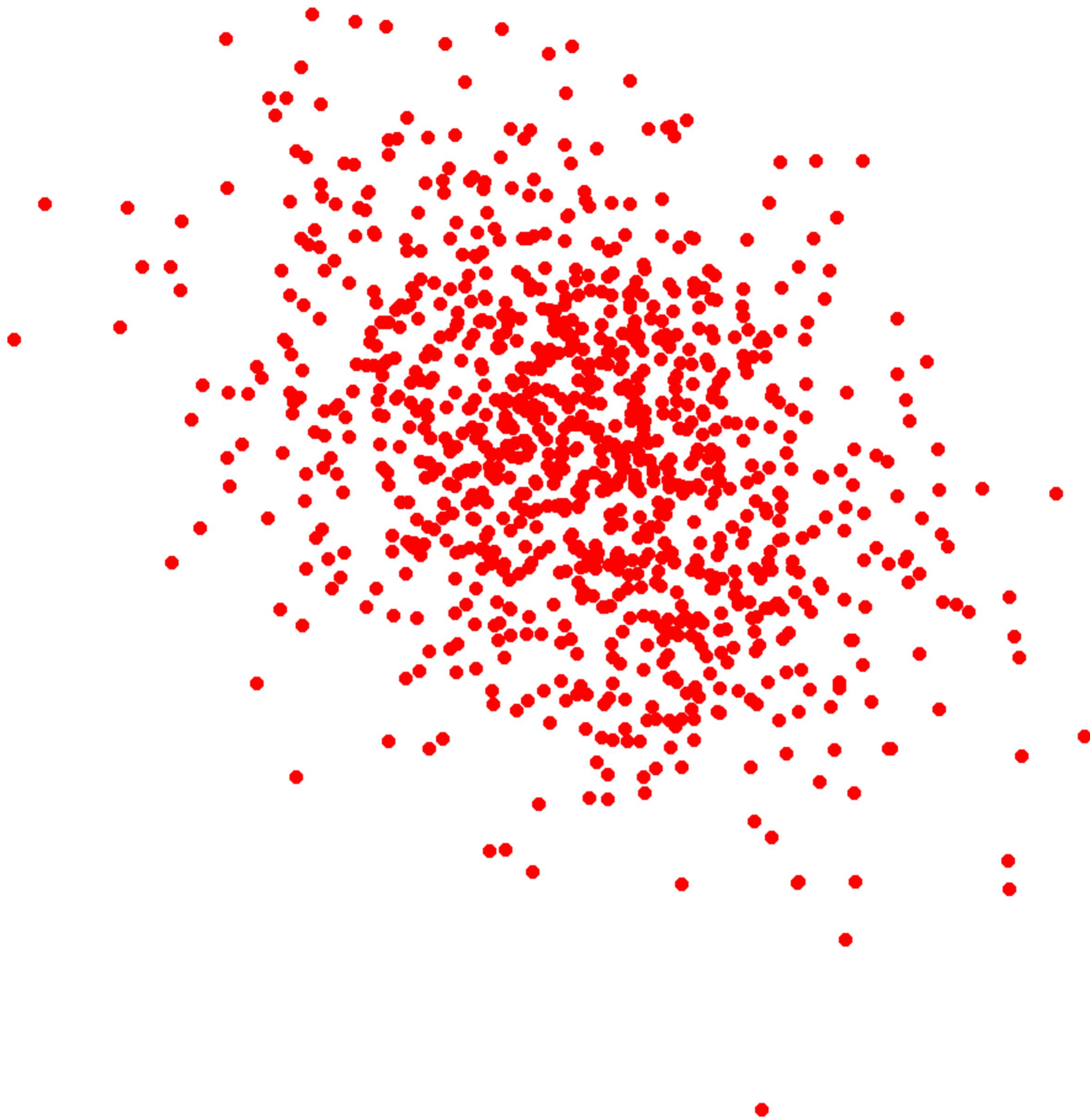
-0.37



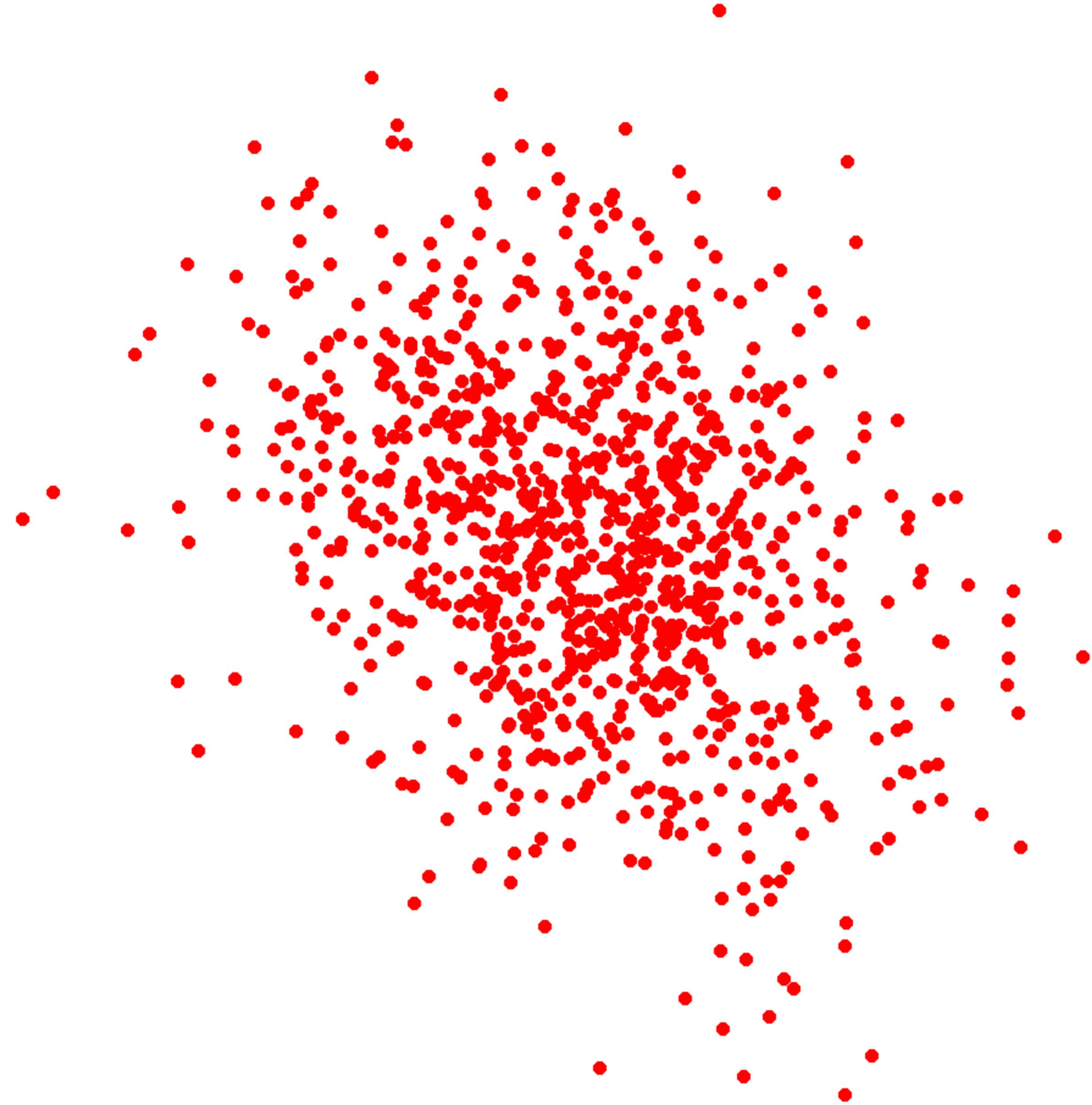
-0.33



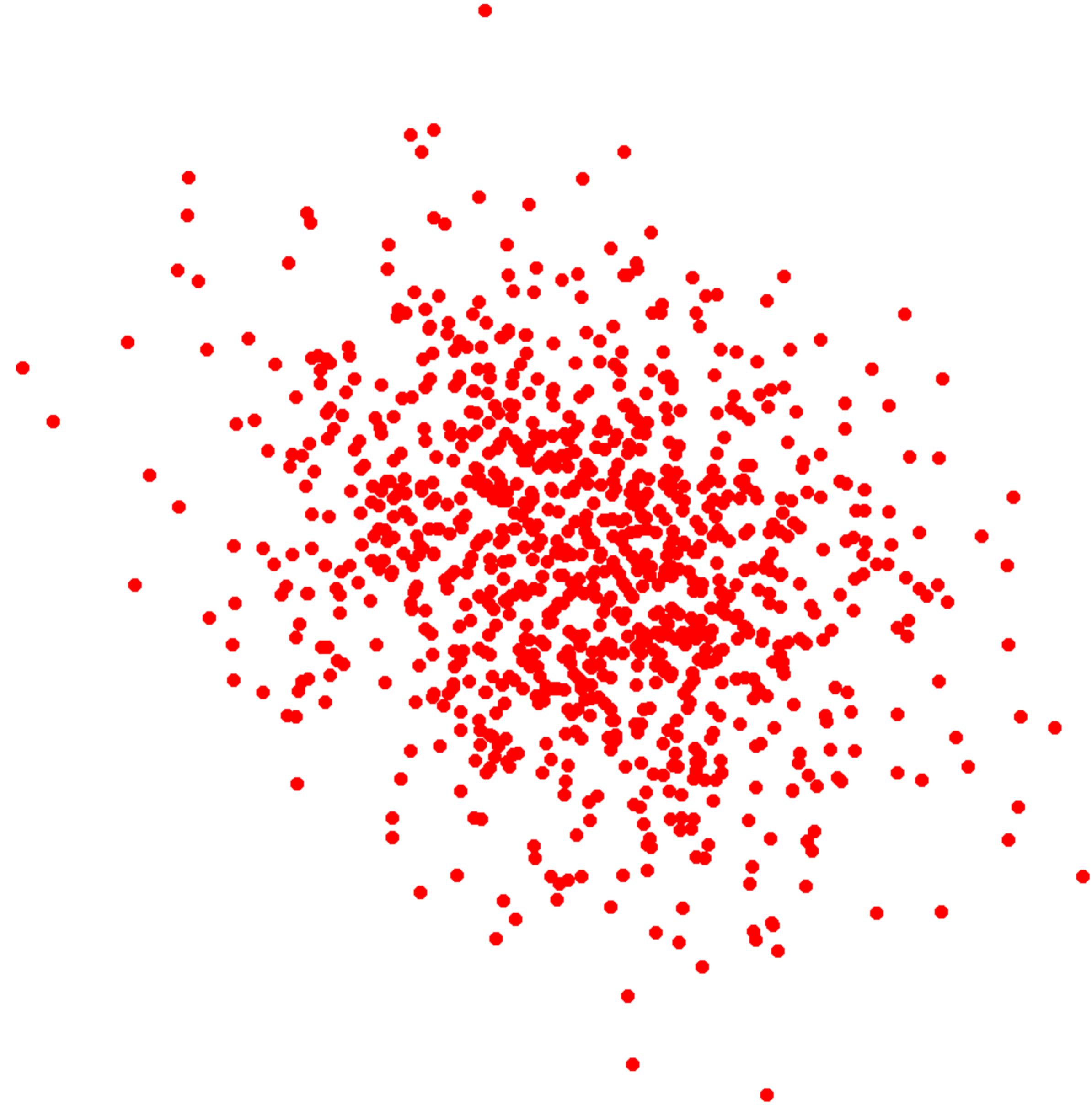
-0.33



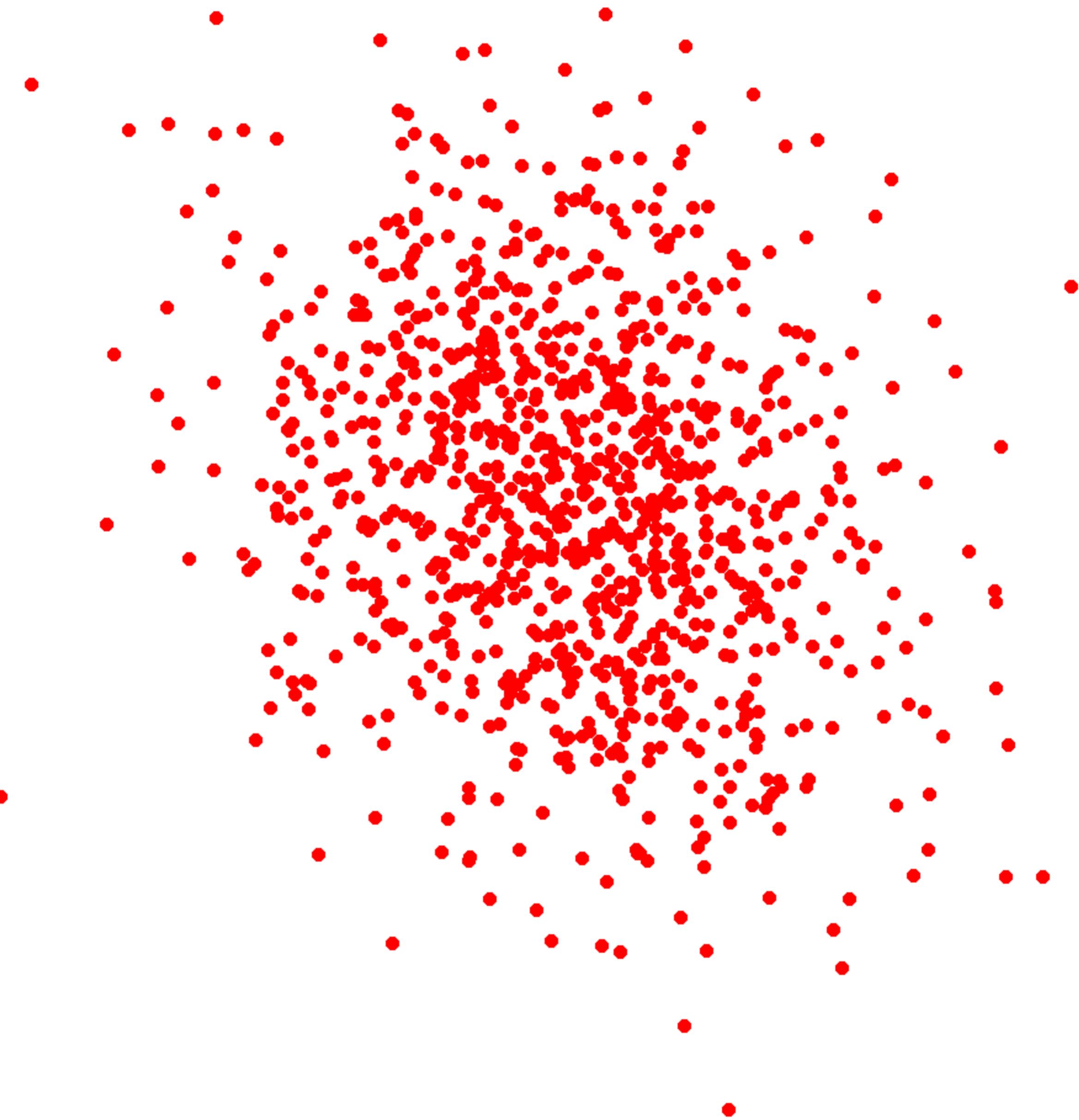
-0.28



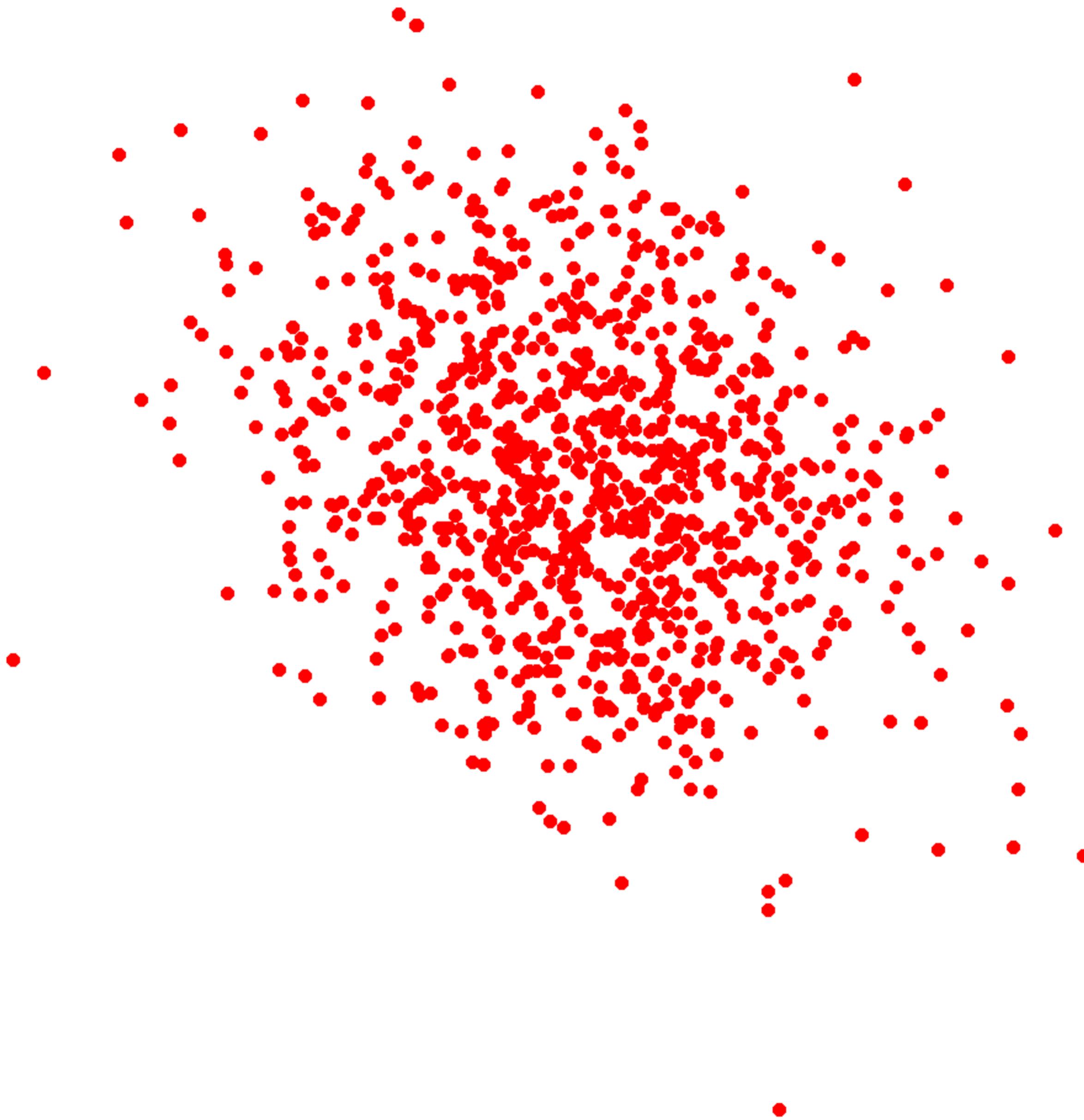
-0.24



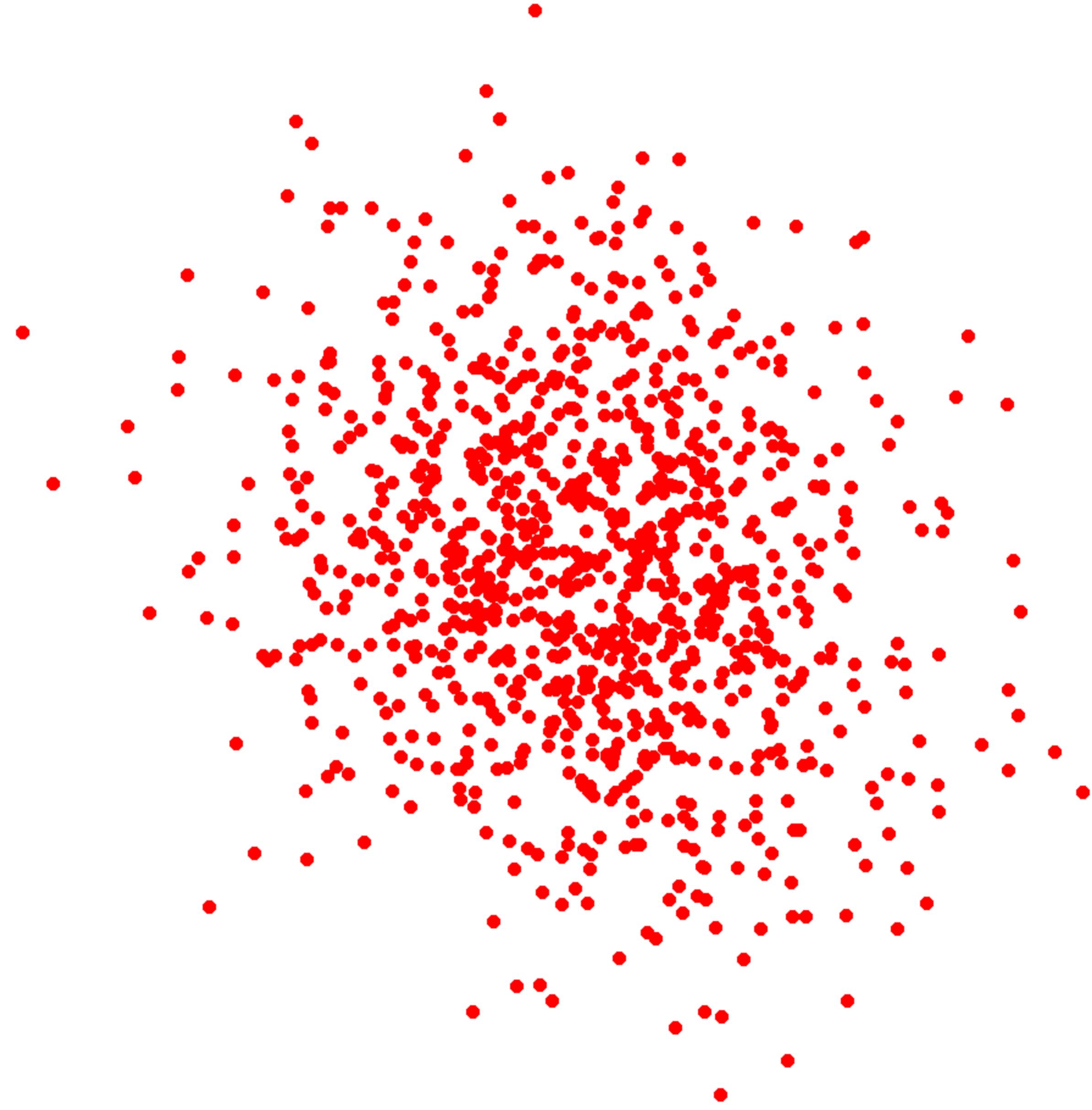
-0.22



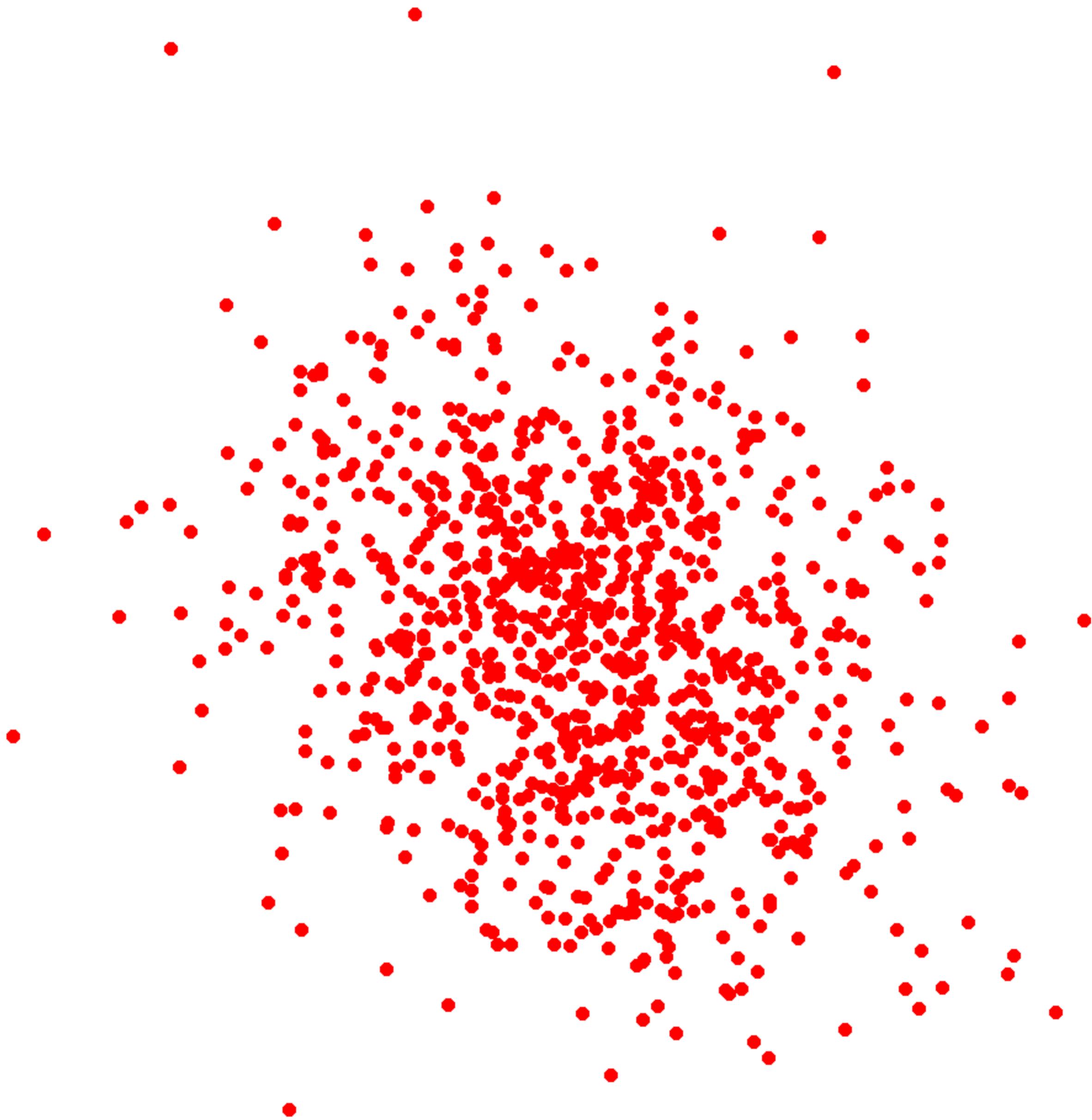
-0.23



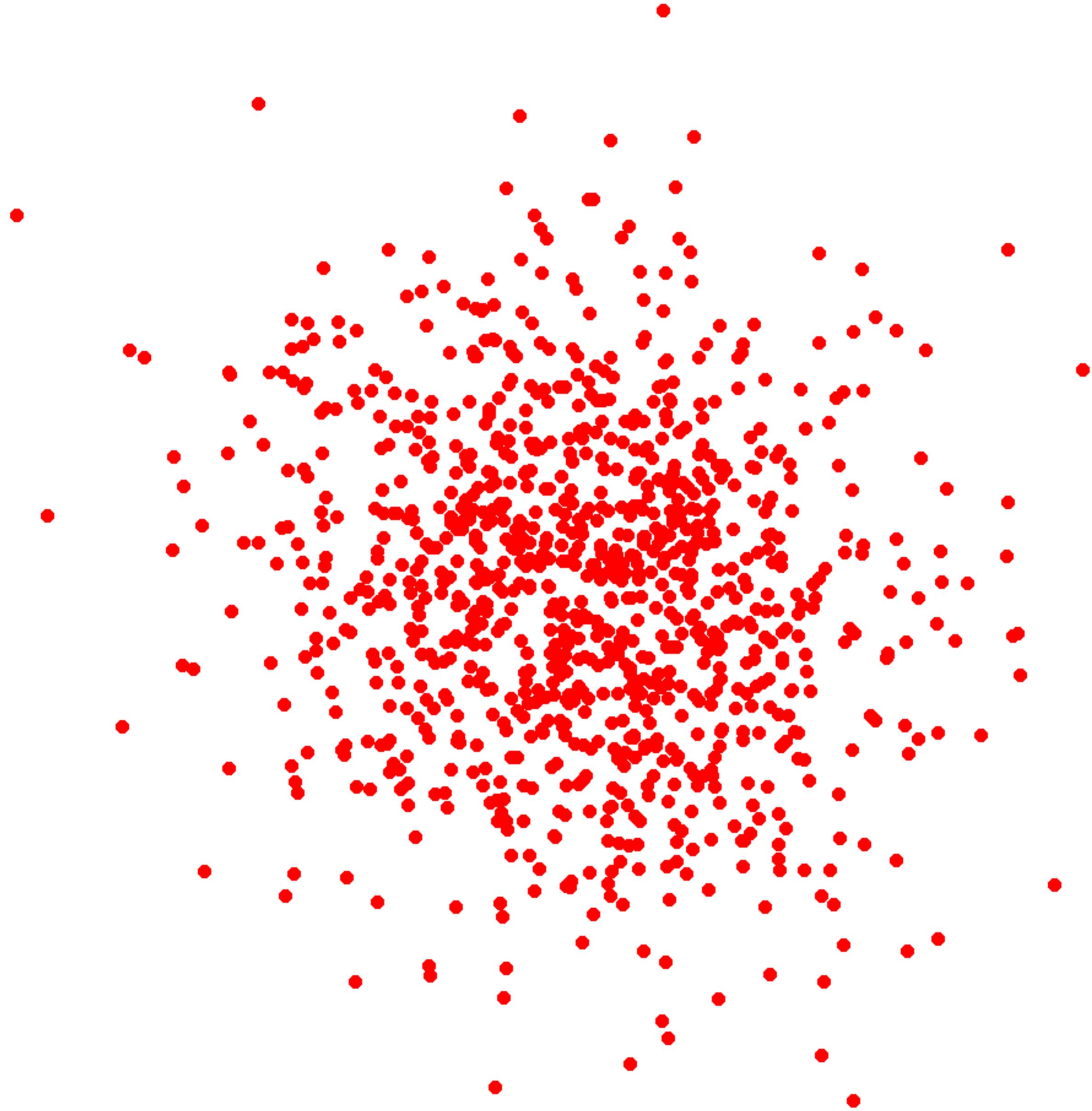
-0.18



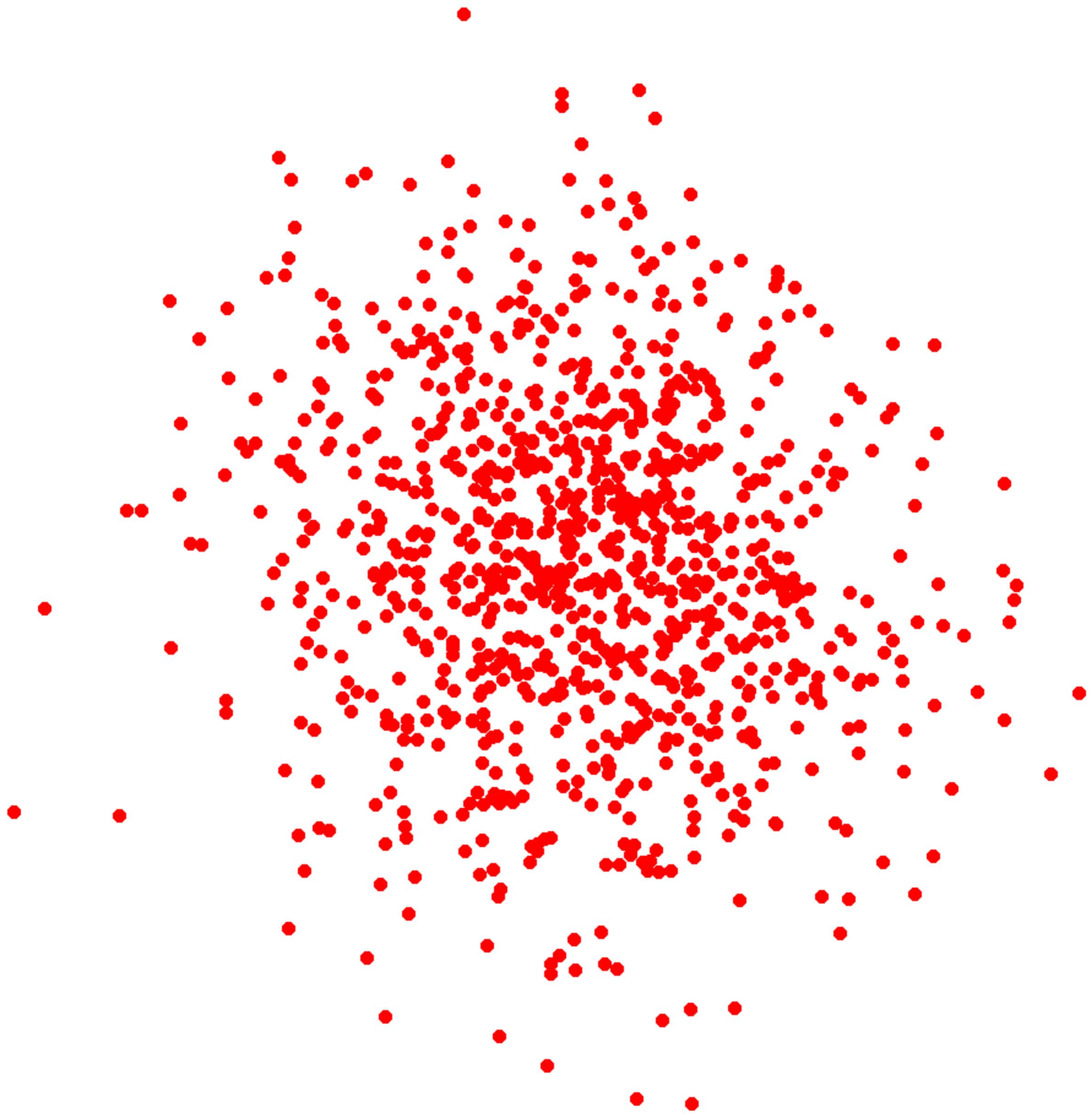
-0.25



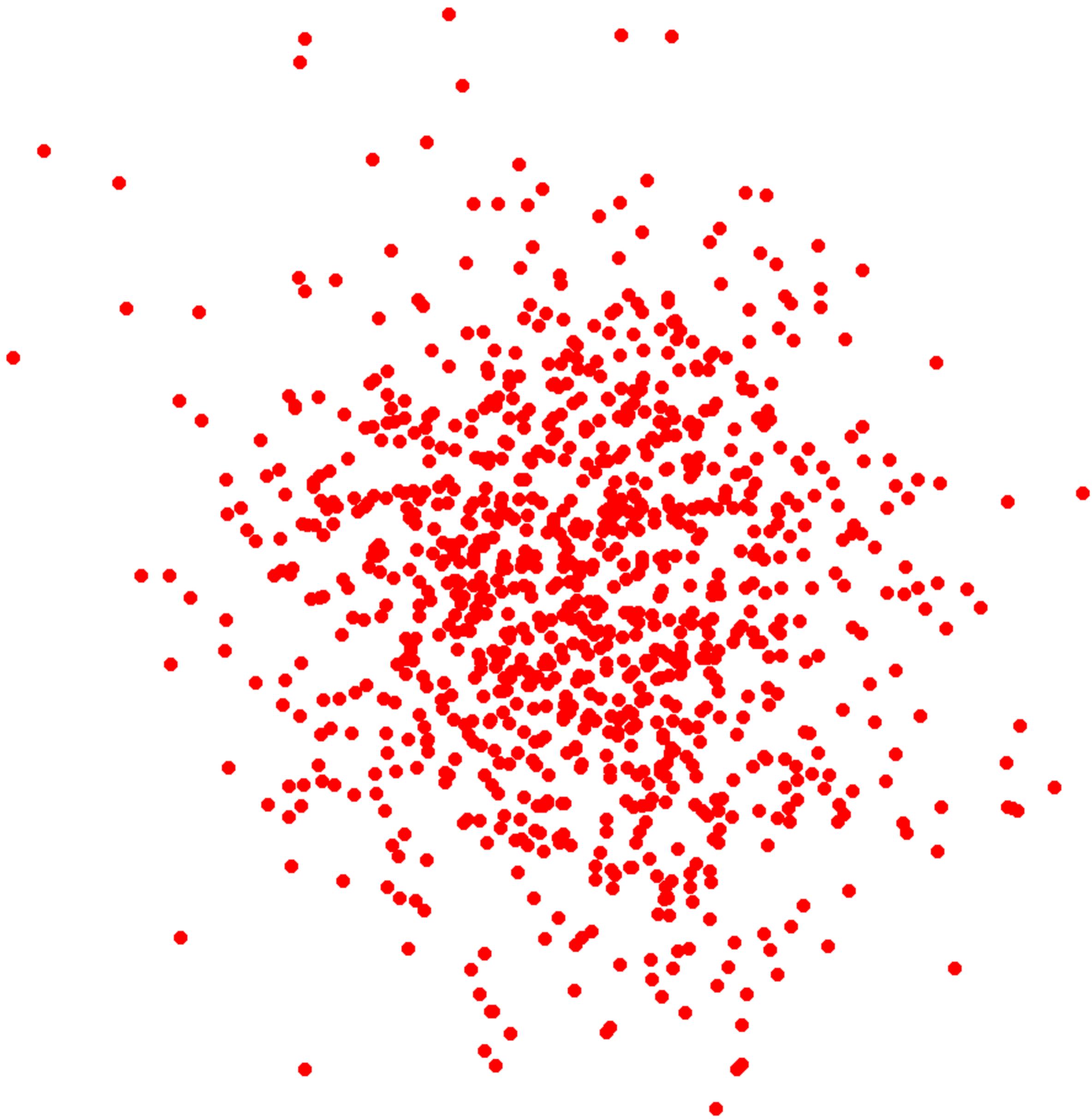
-0.12



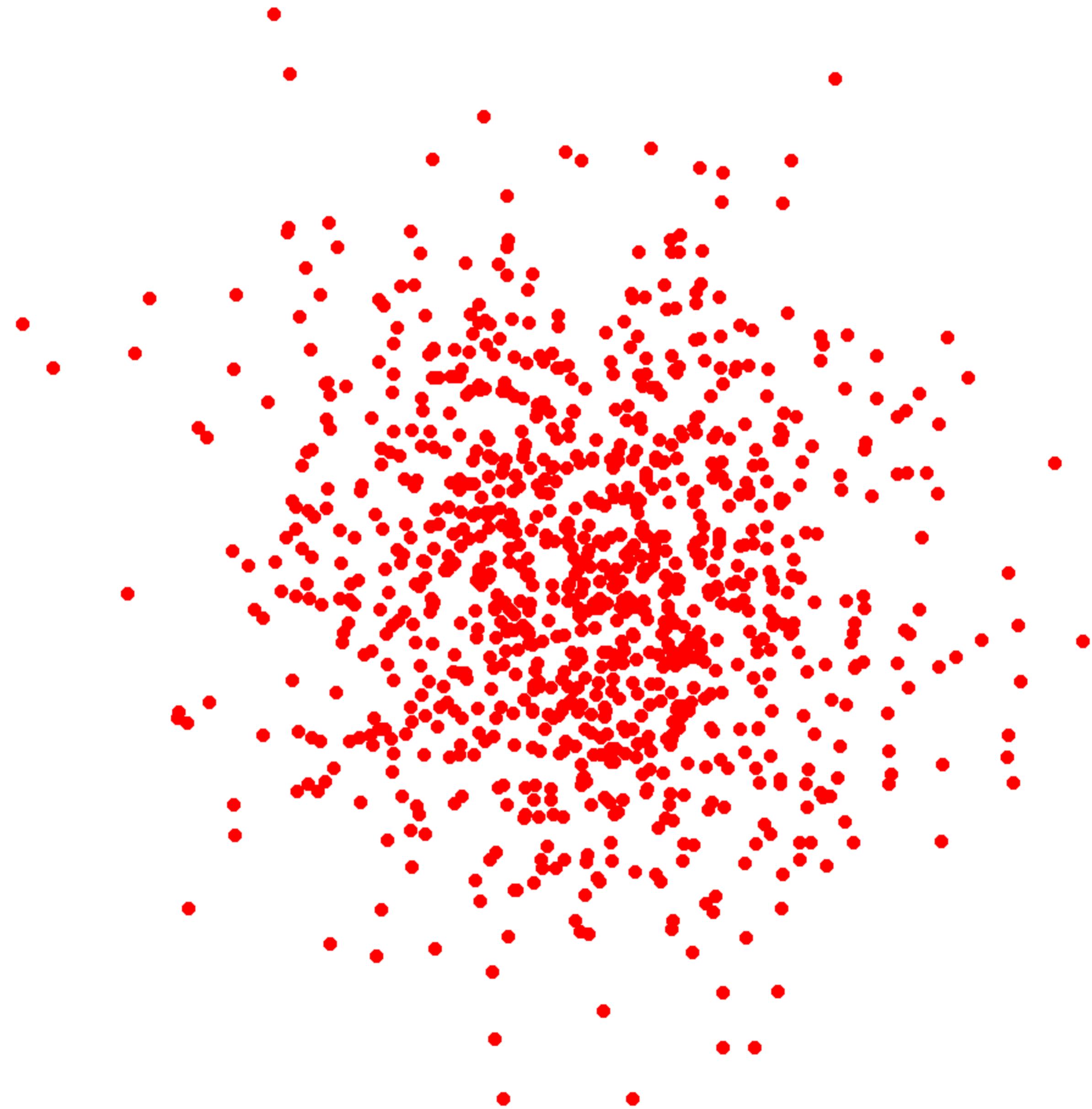
-0.11



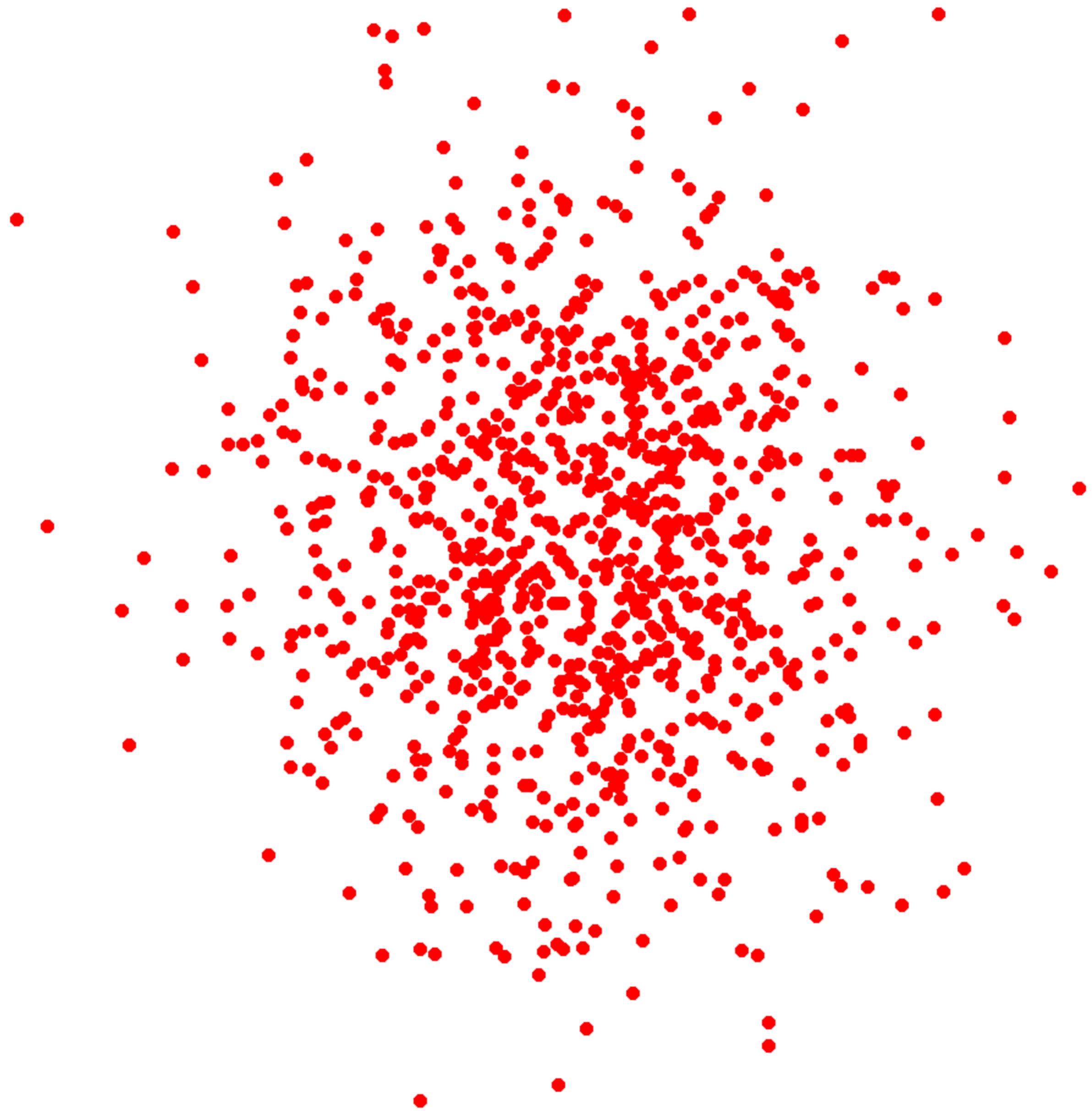
-0.07



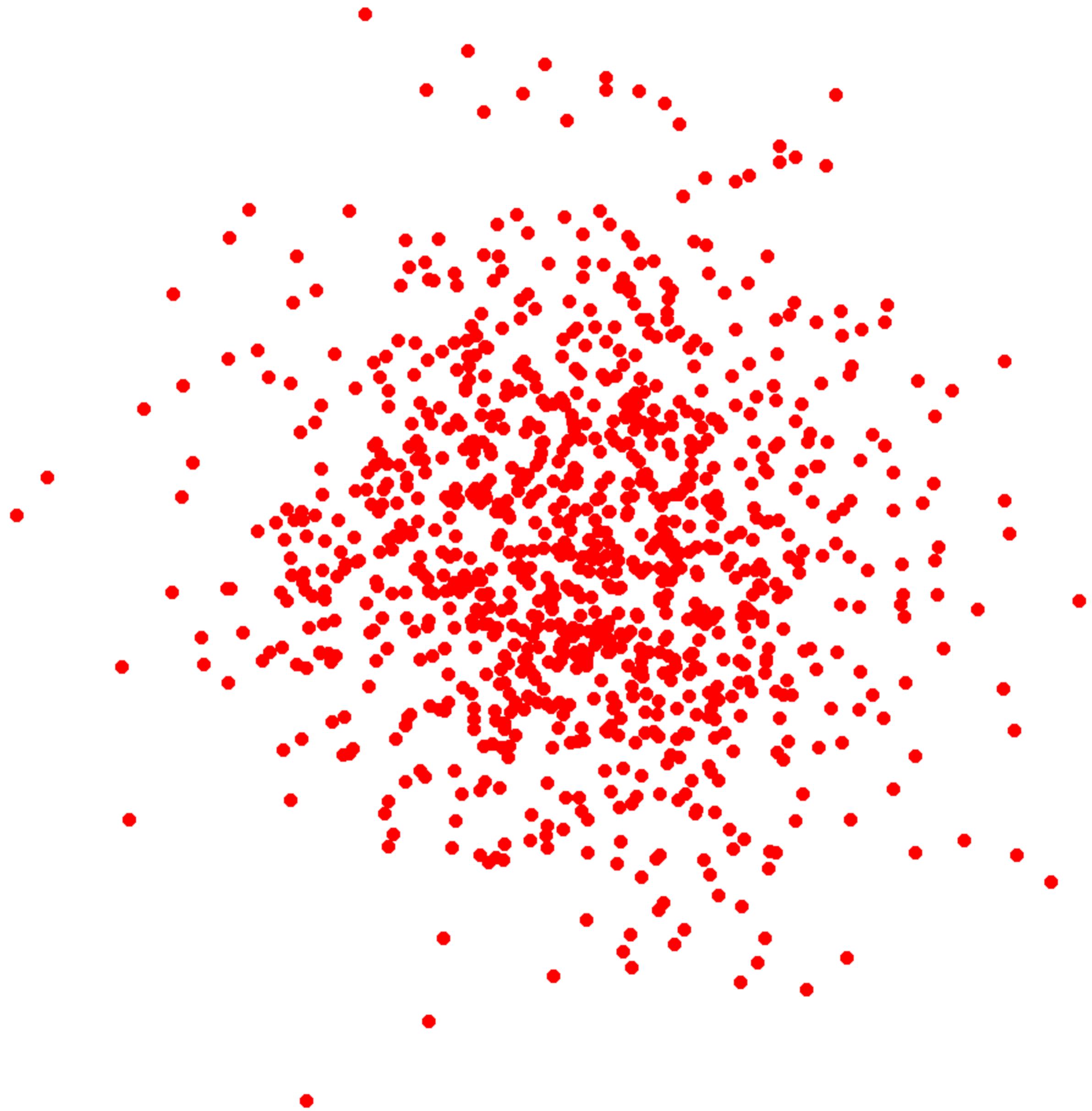
-0.08



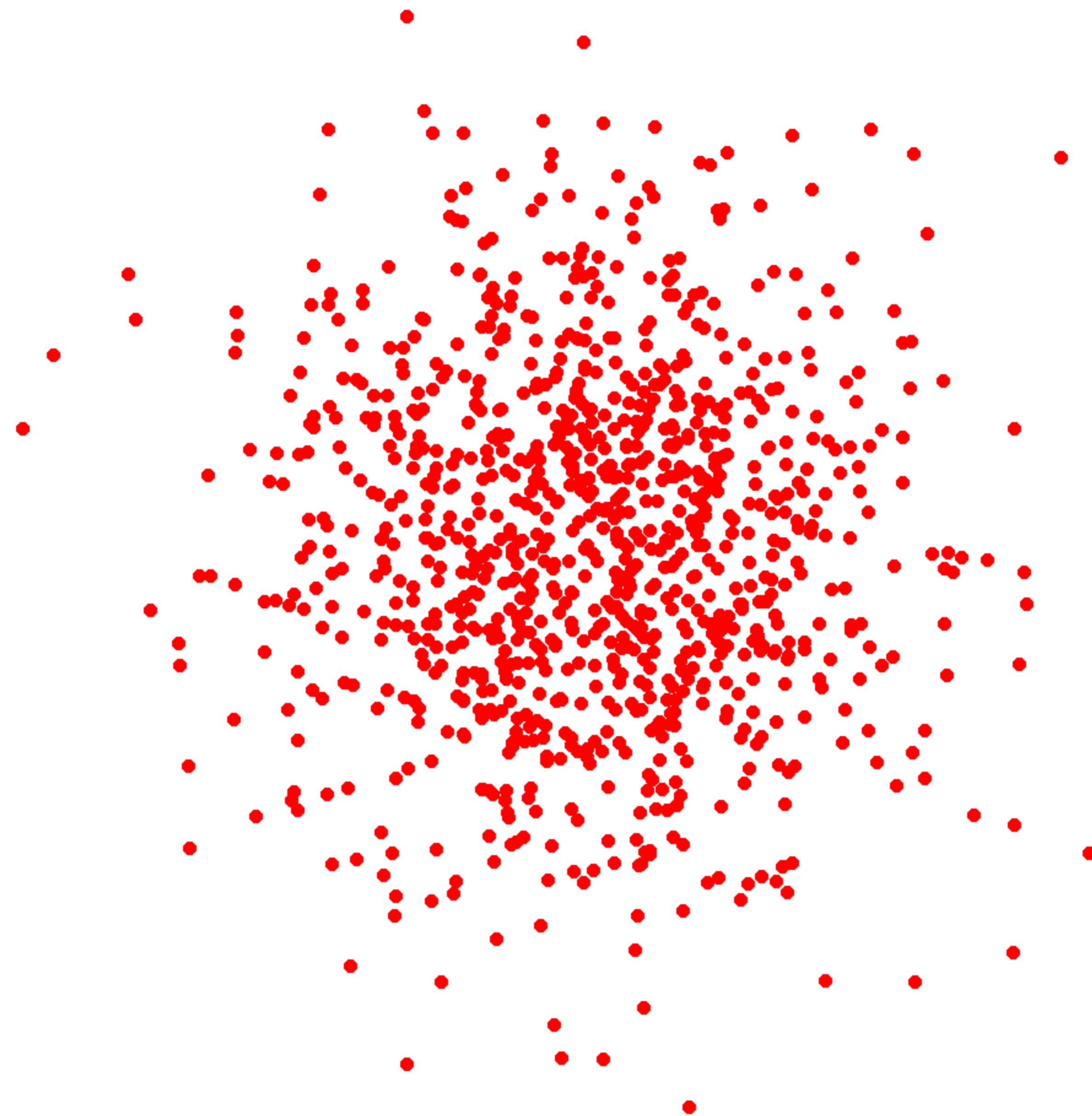
-0.02



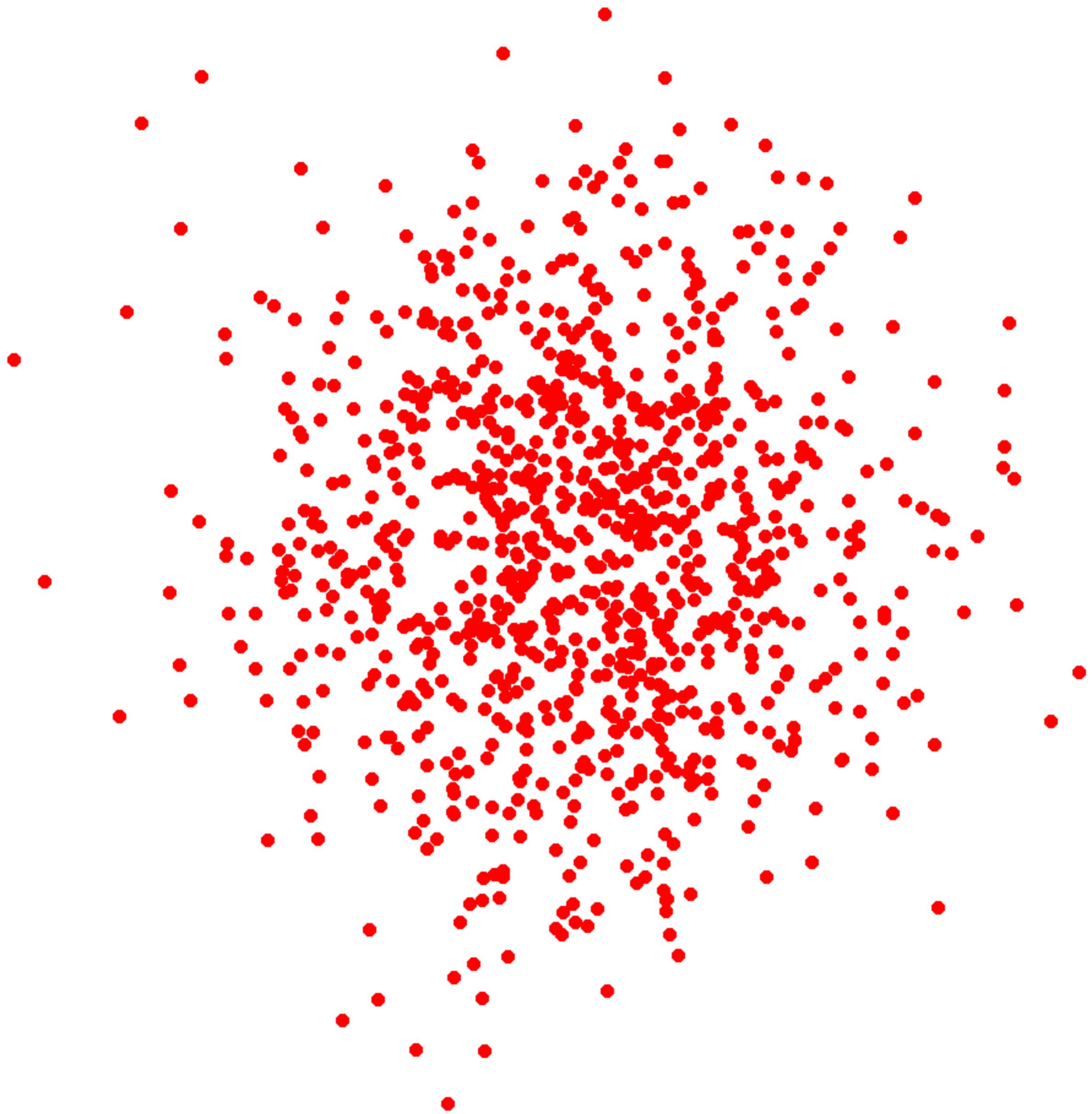
-0.06



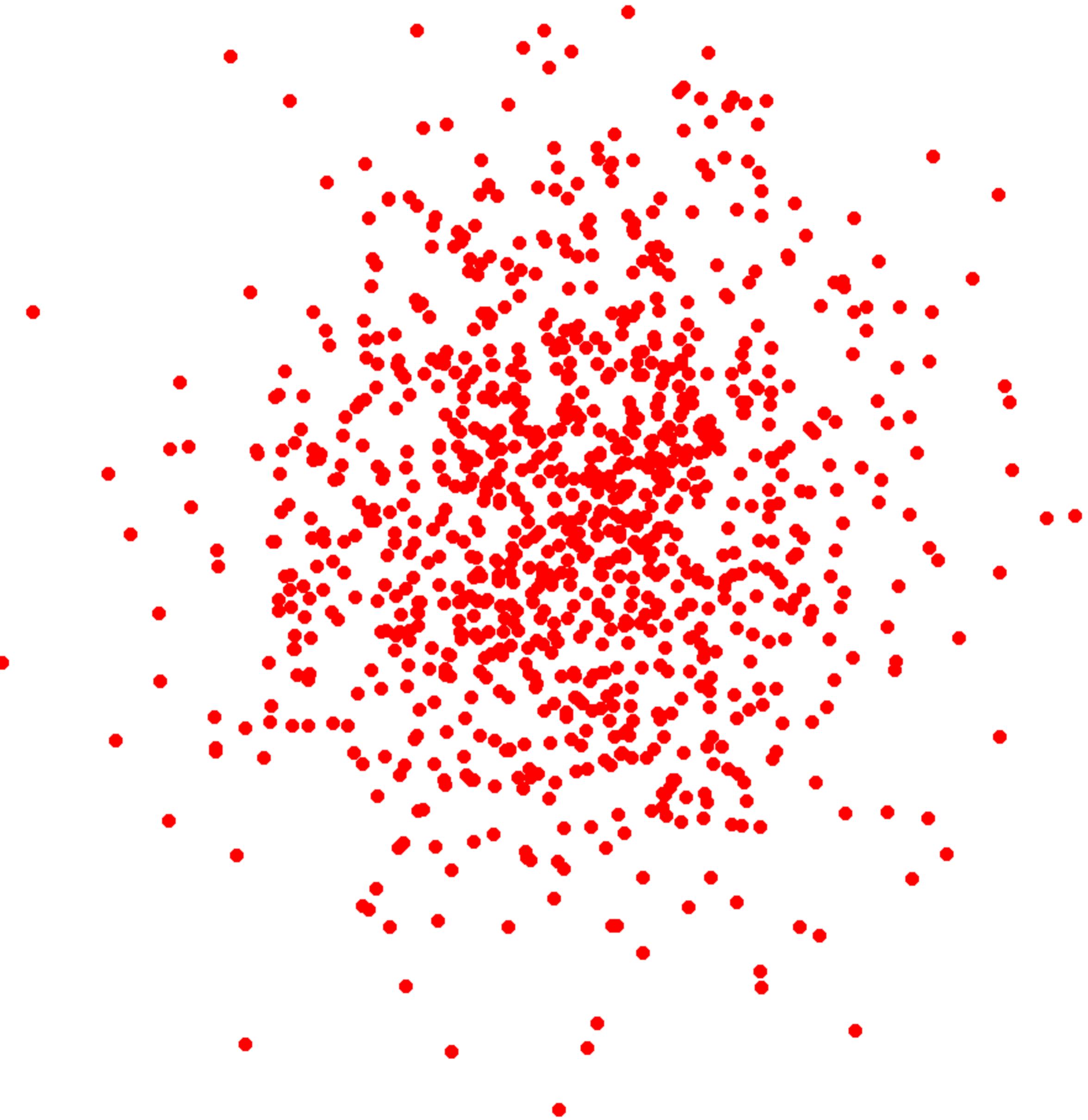
-0.02



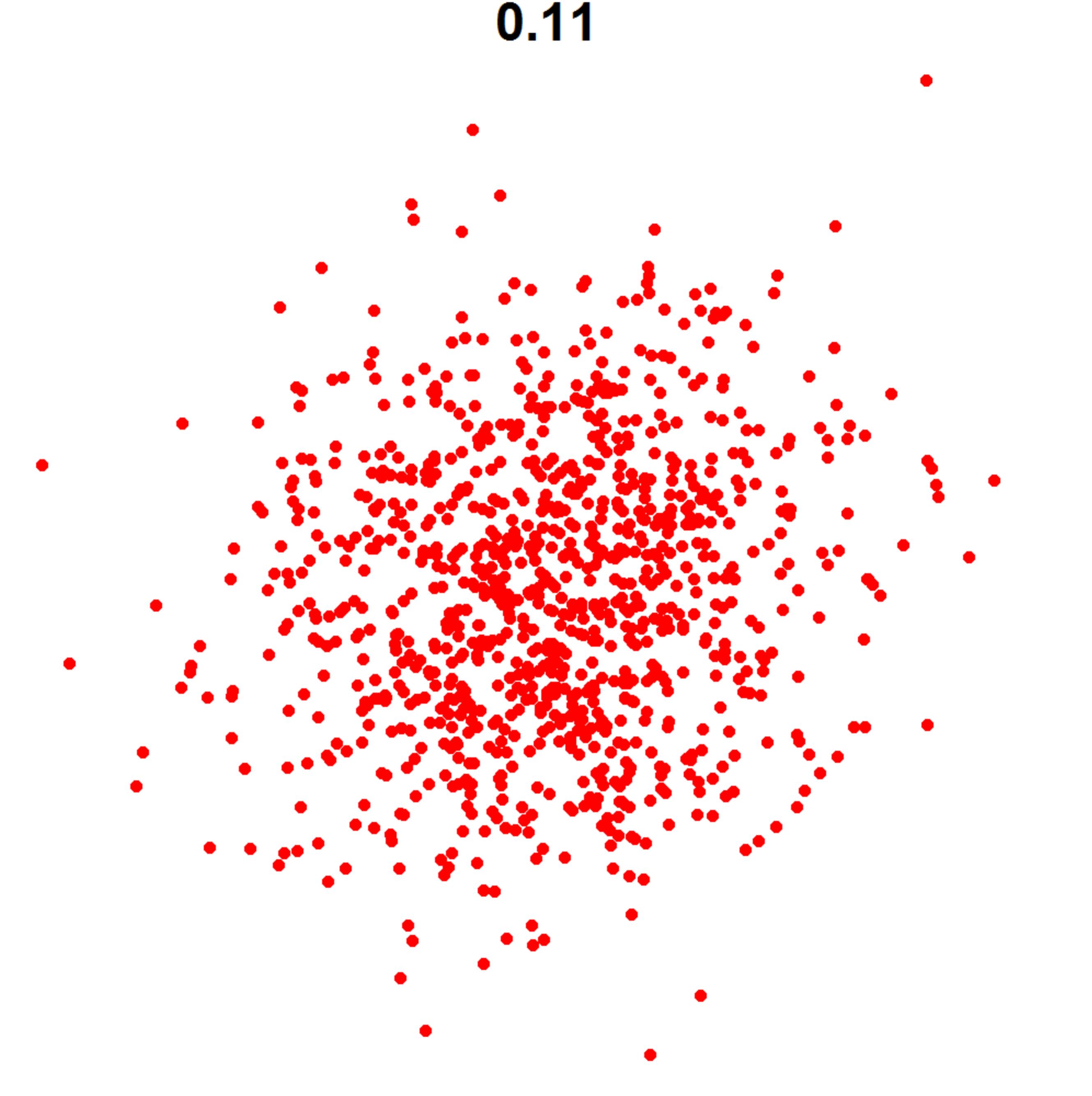
0.02



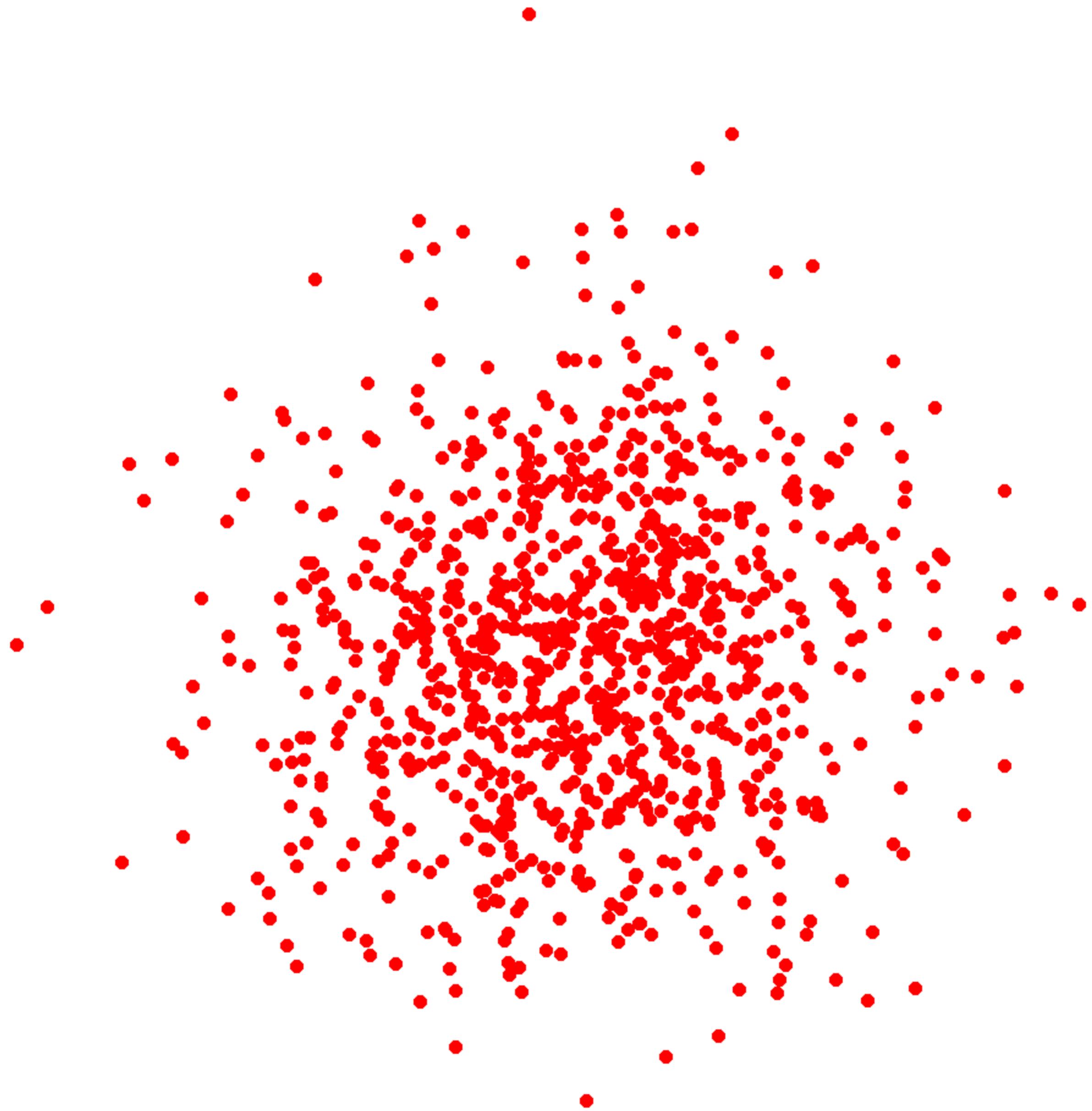
0.06



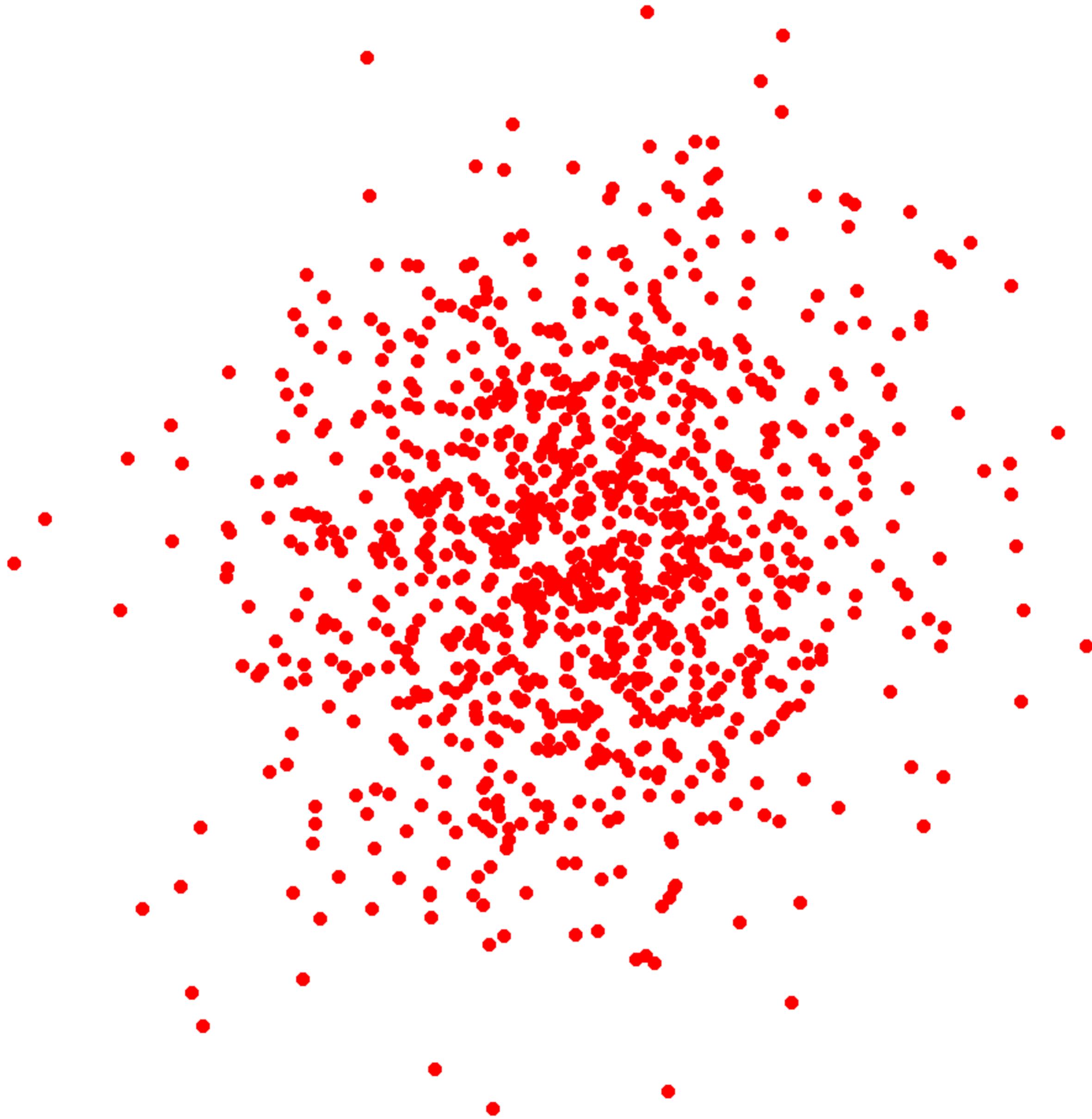
0.11



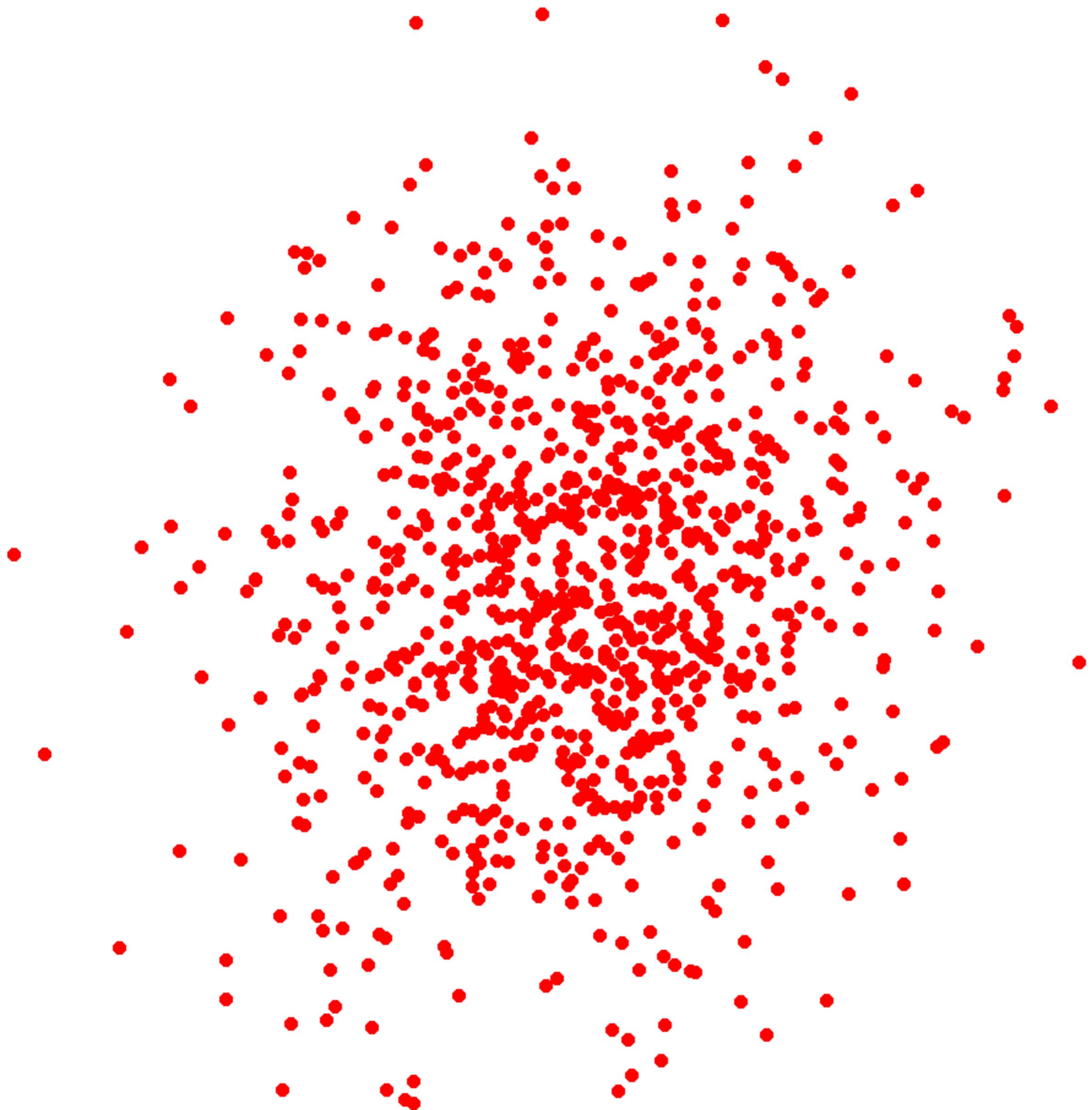
0.07



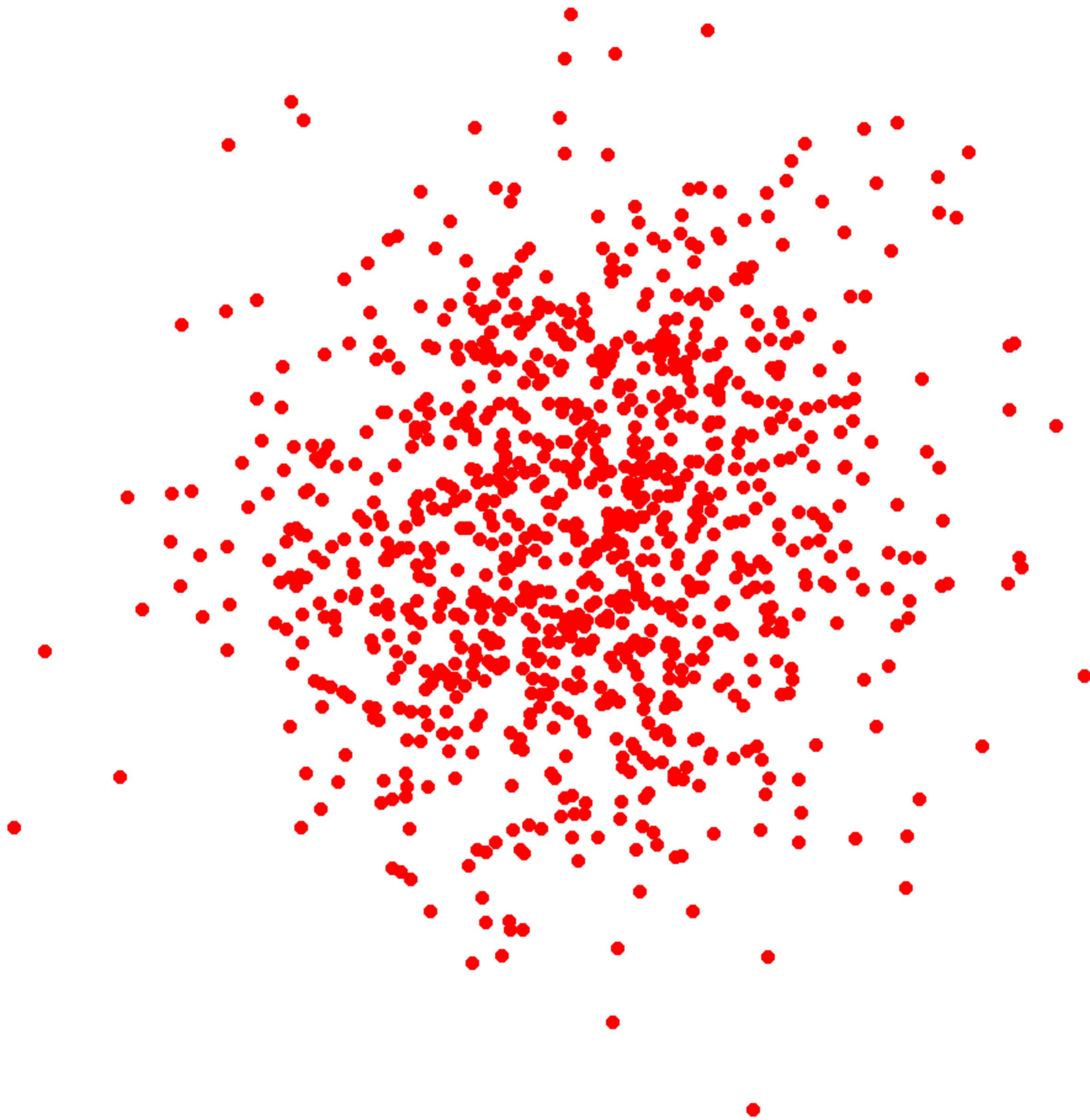
0.1



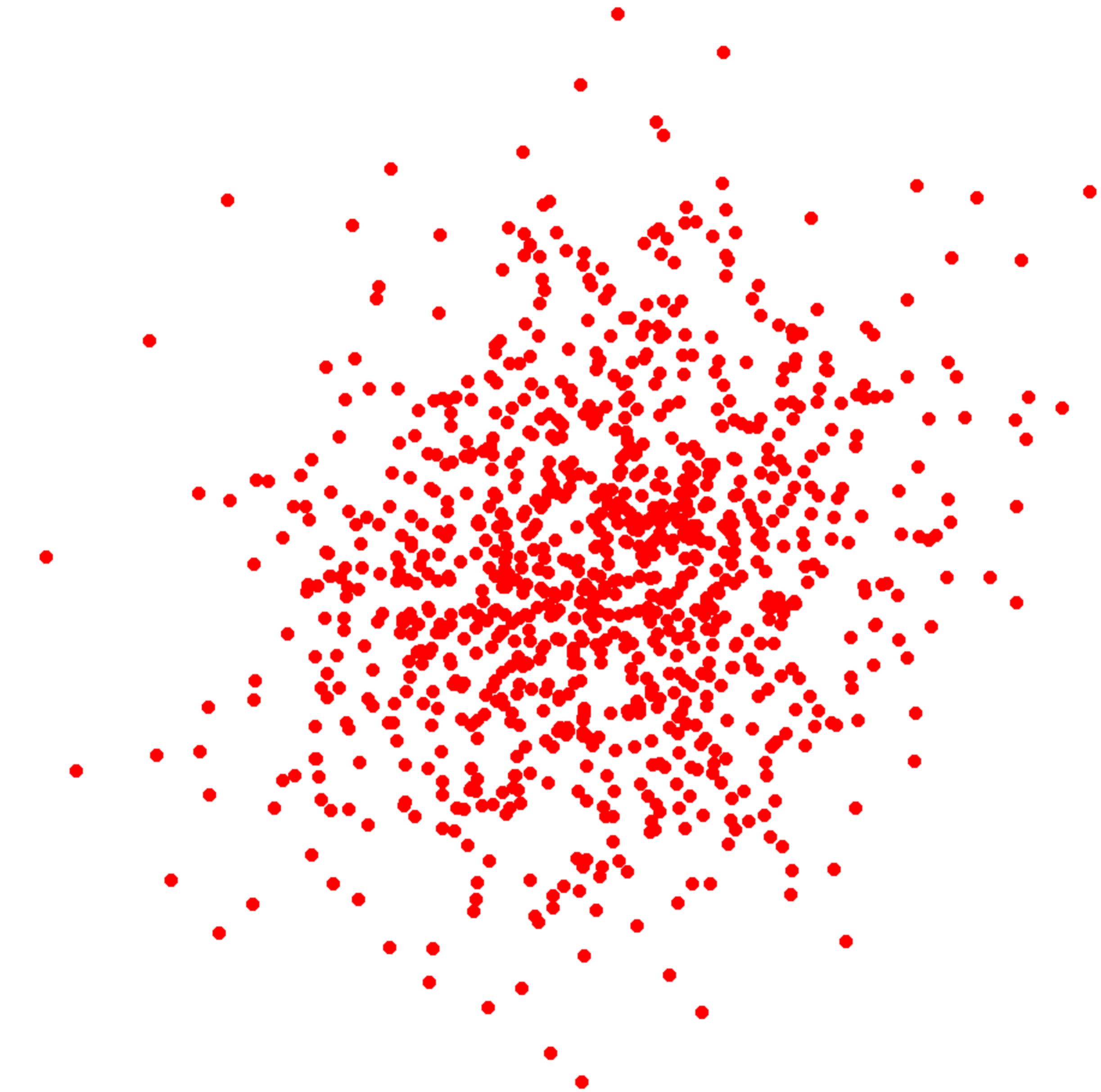
0.14



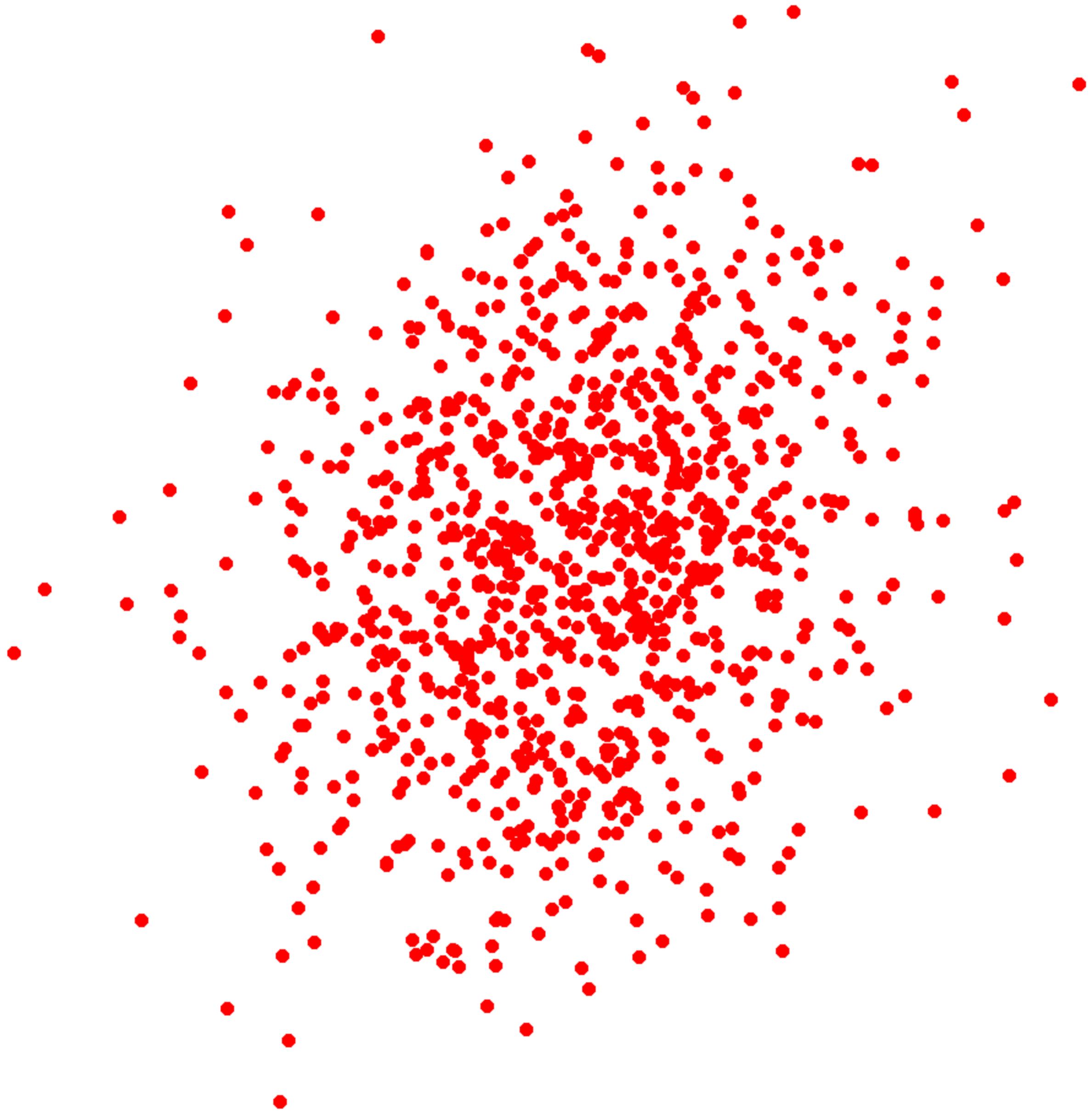
0.12



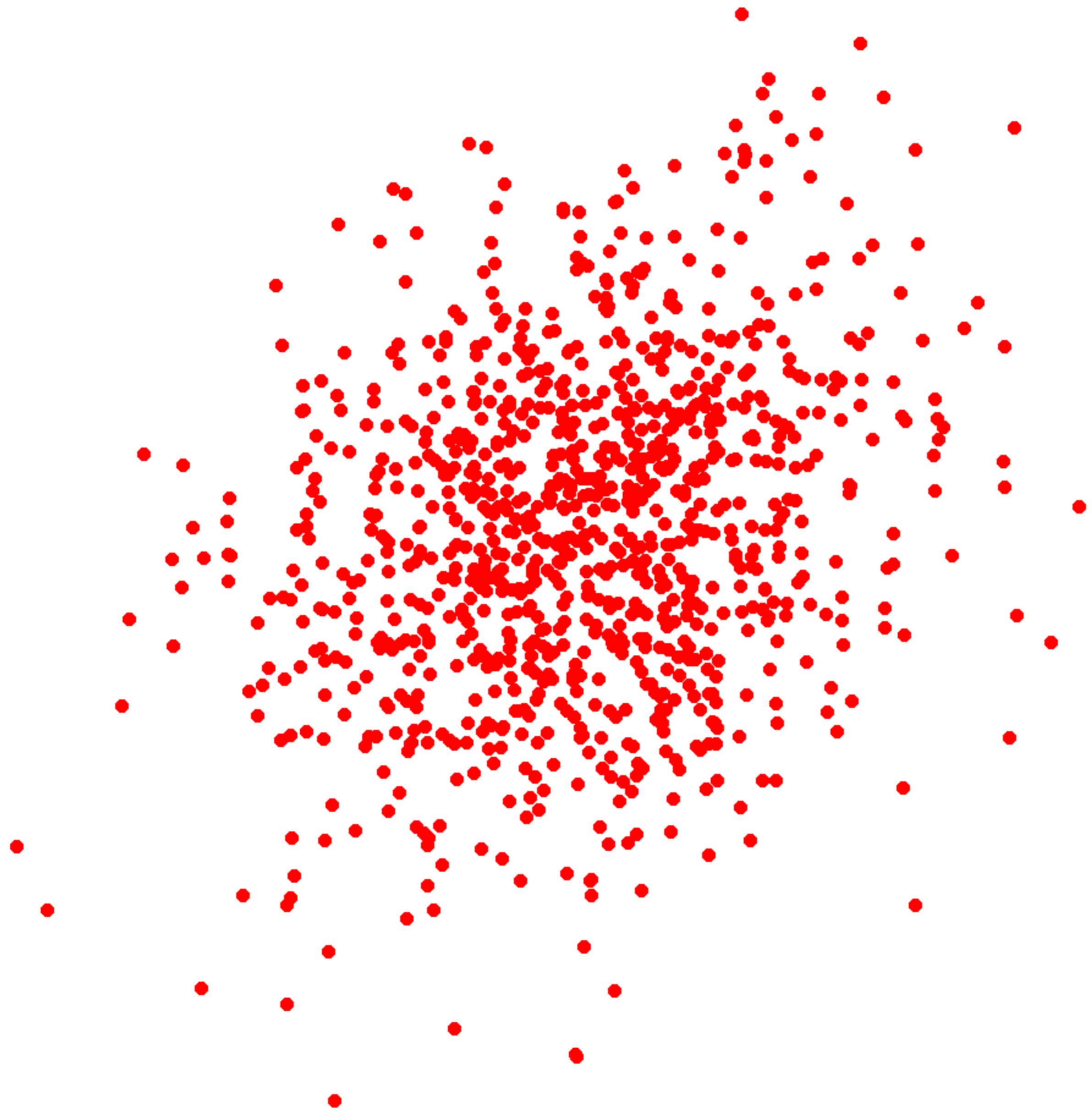
0.16



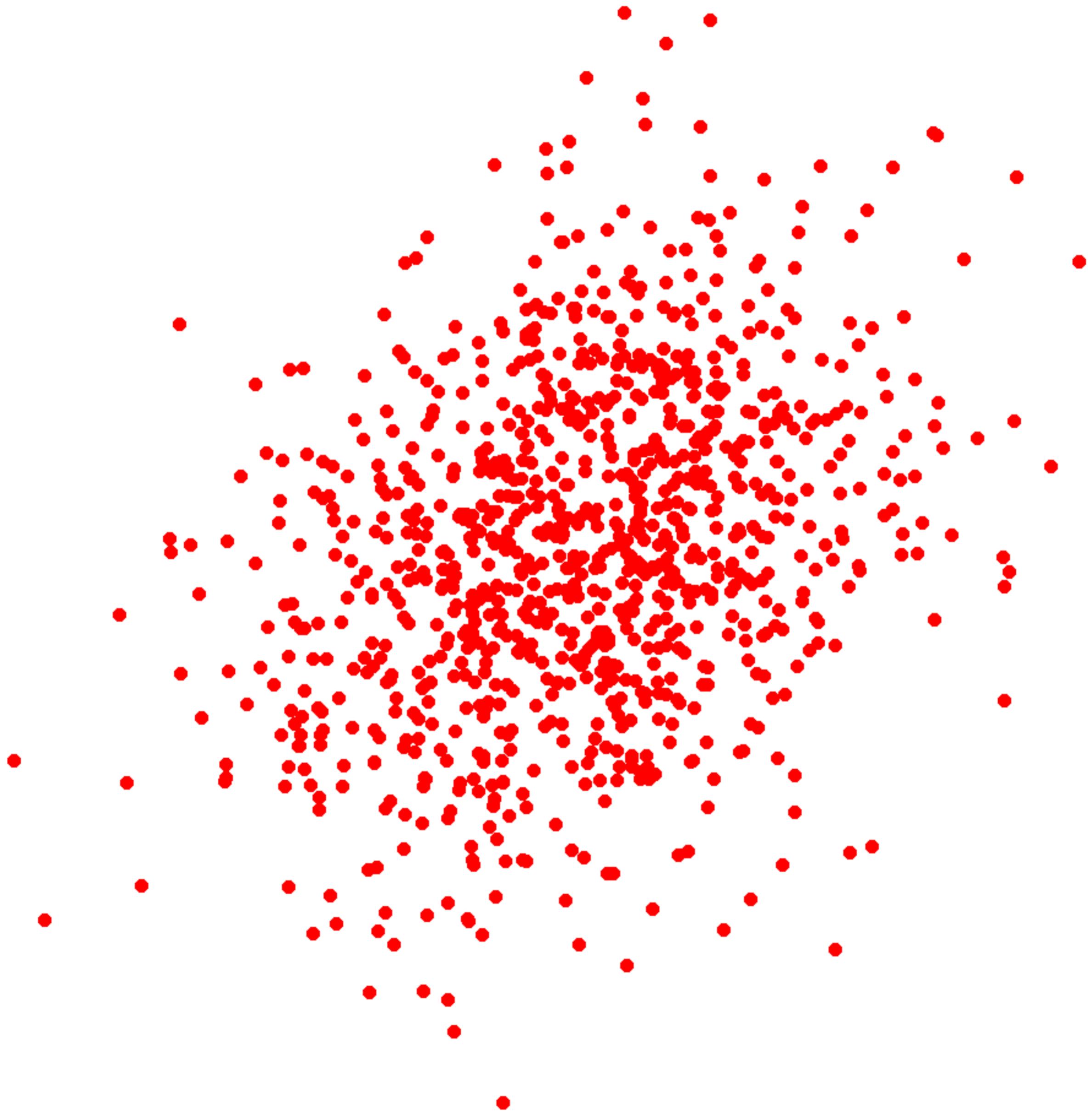
0.21



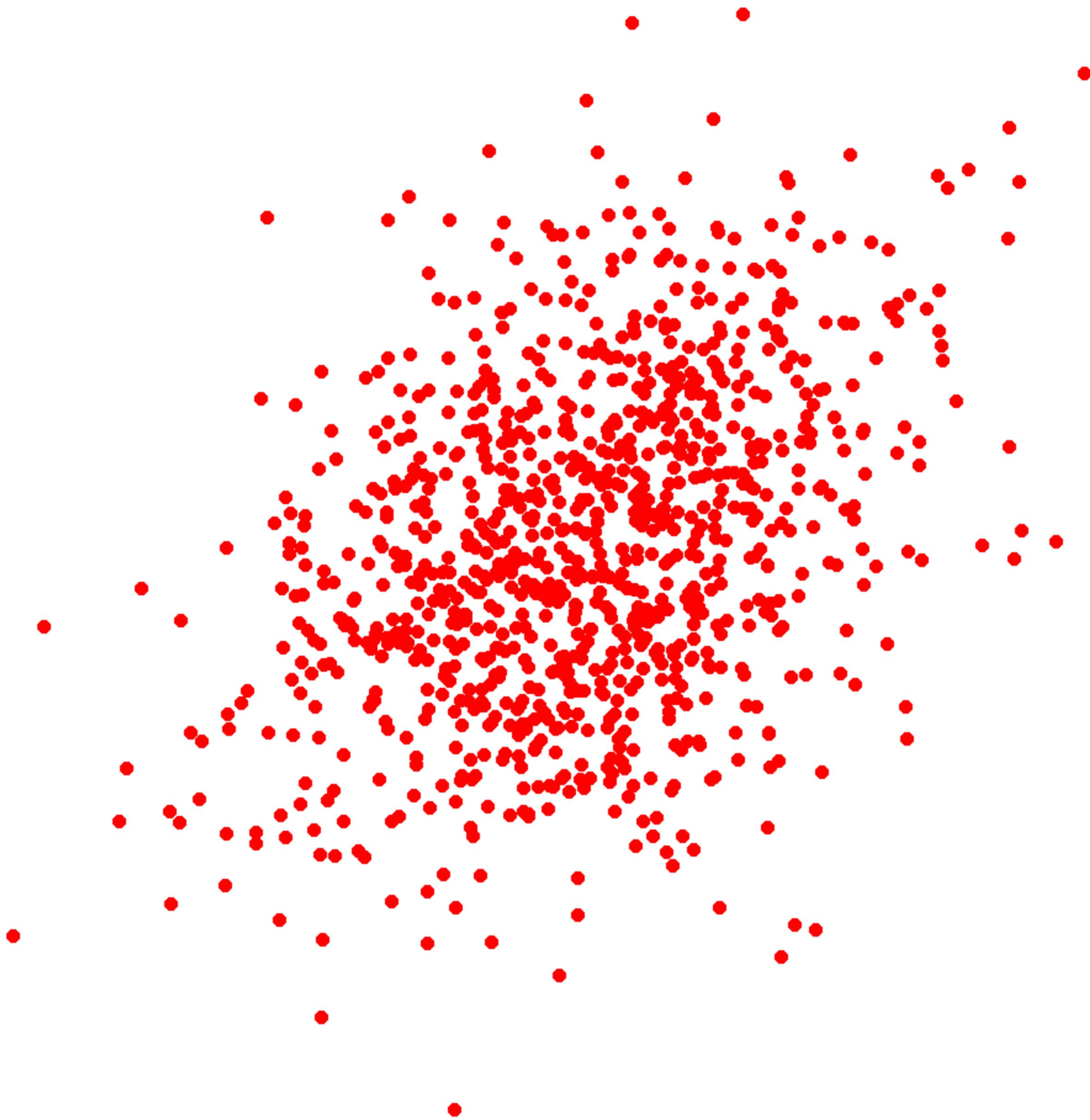
0.22



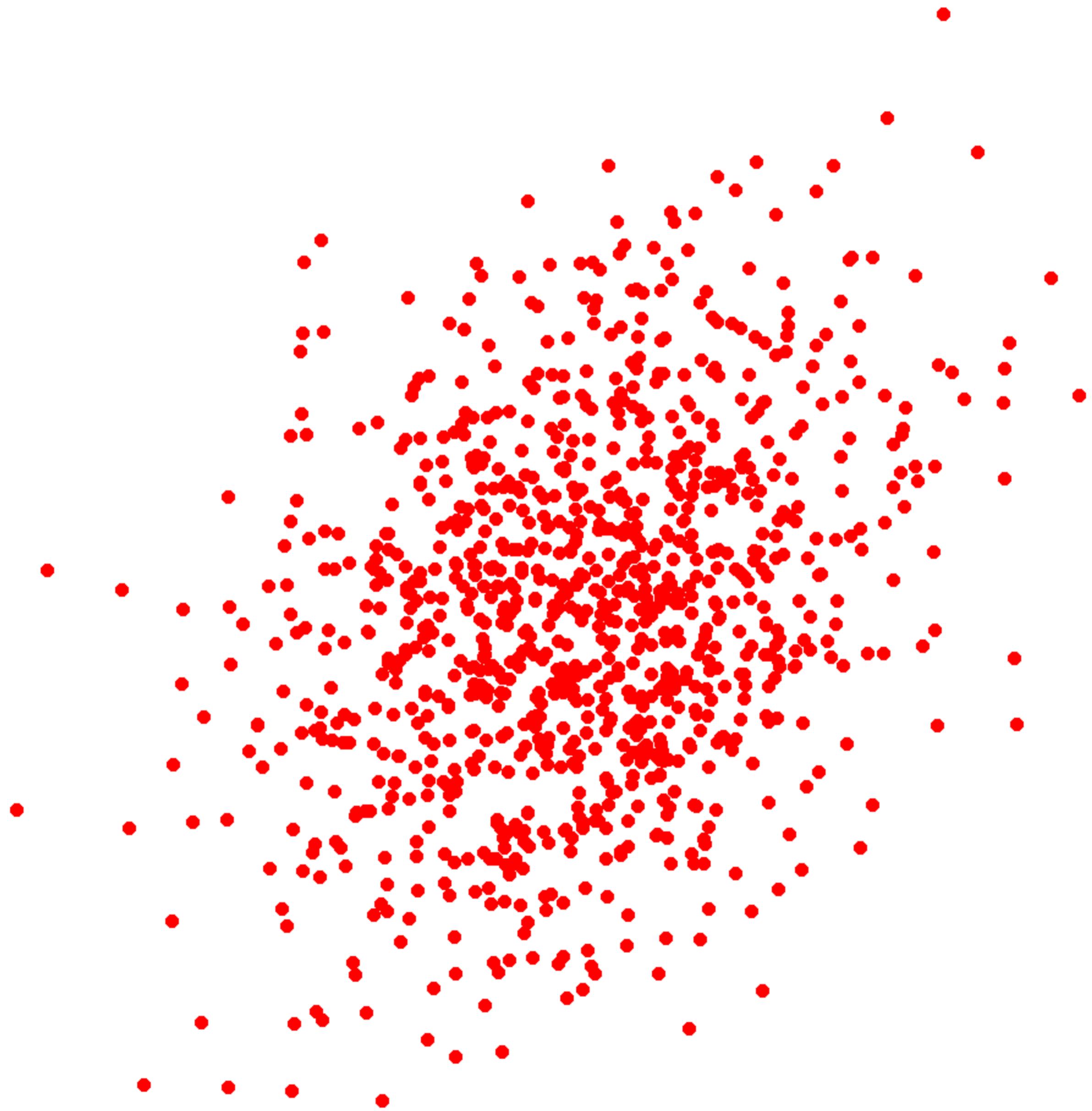
0.31



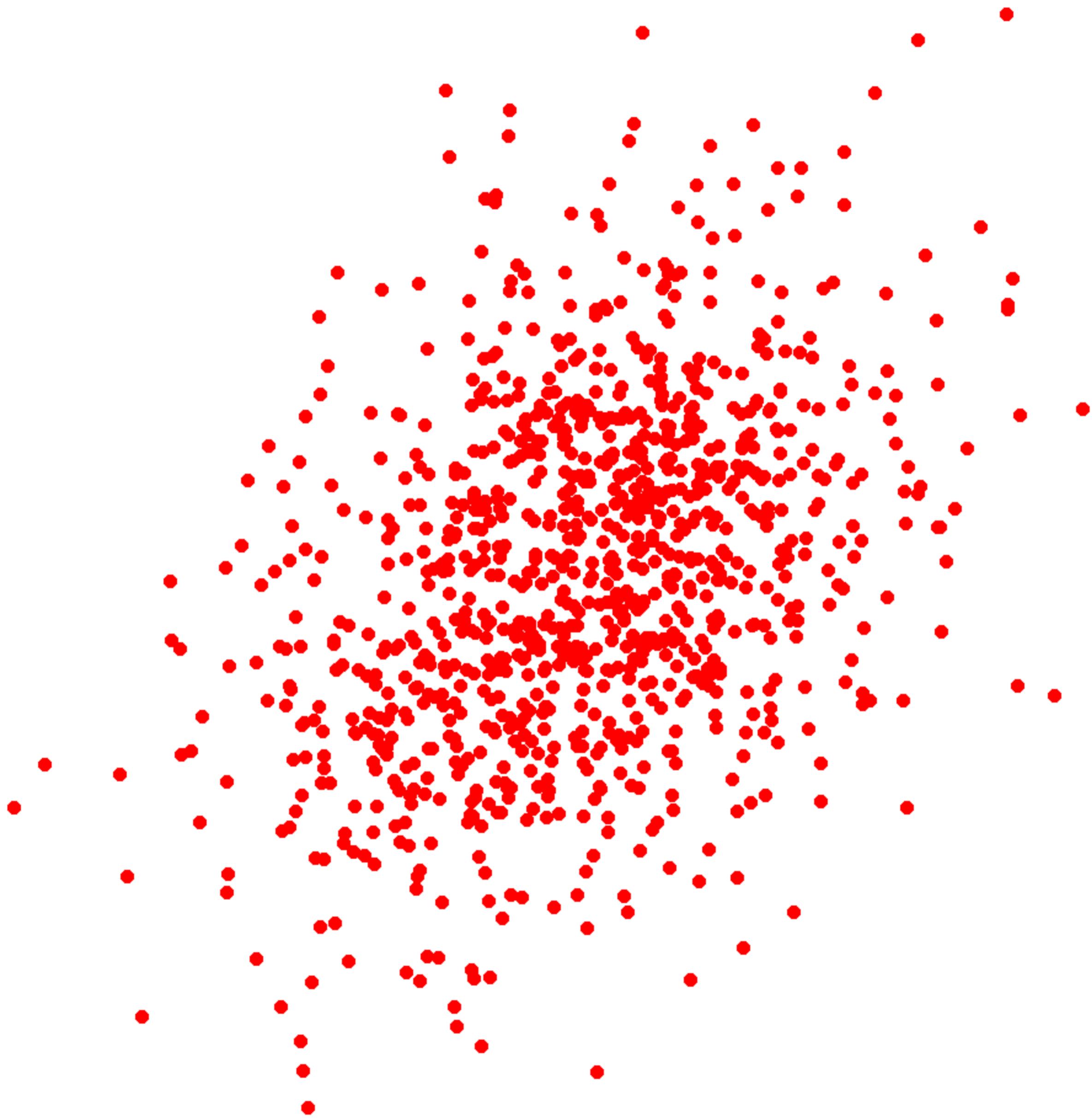
0.33



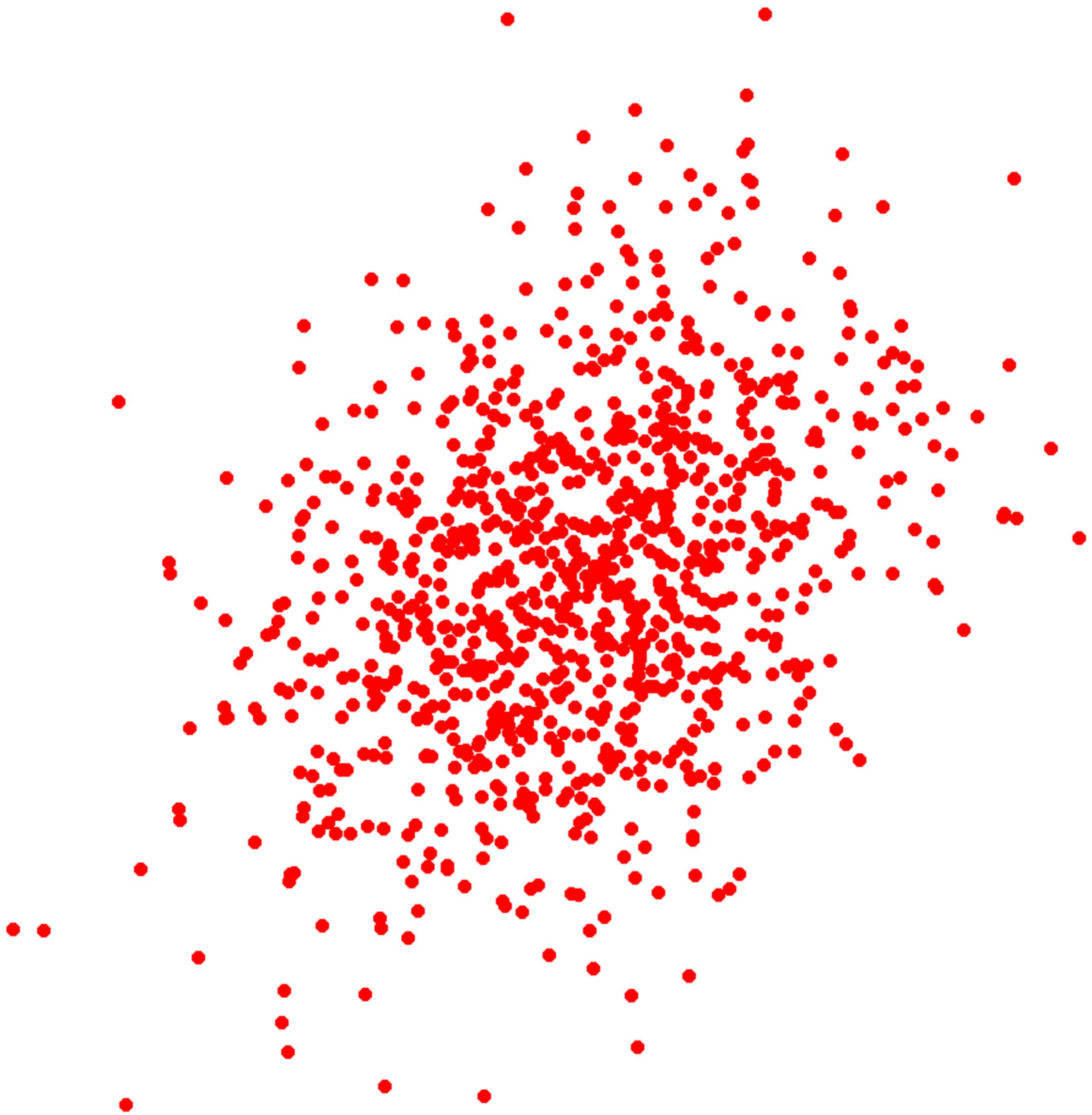
0.27



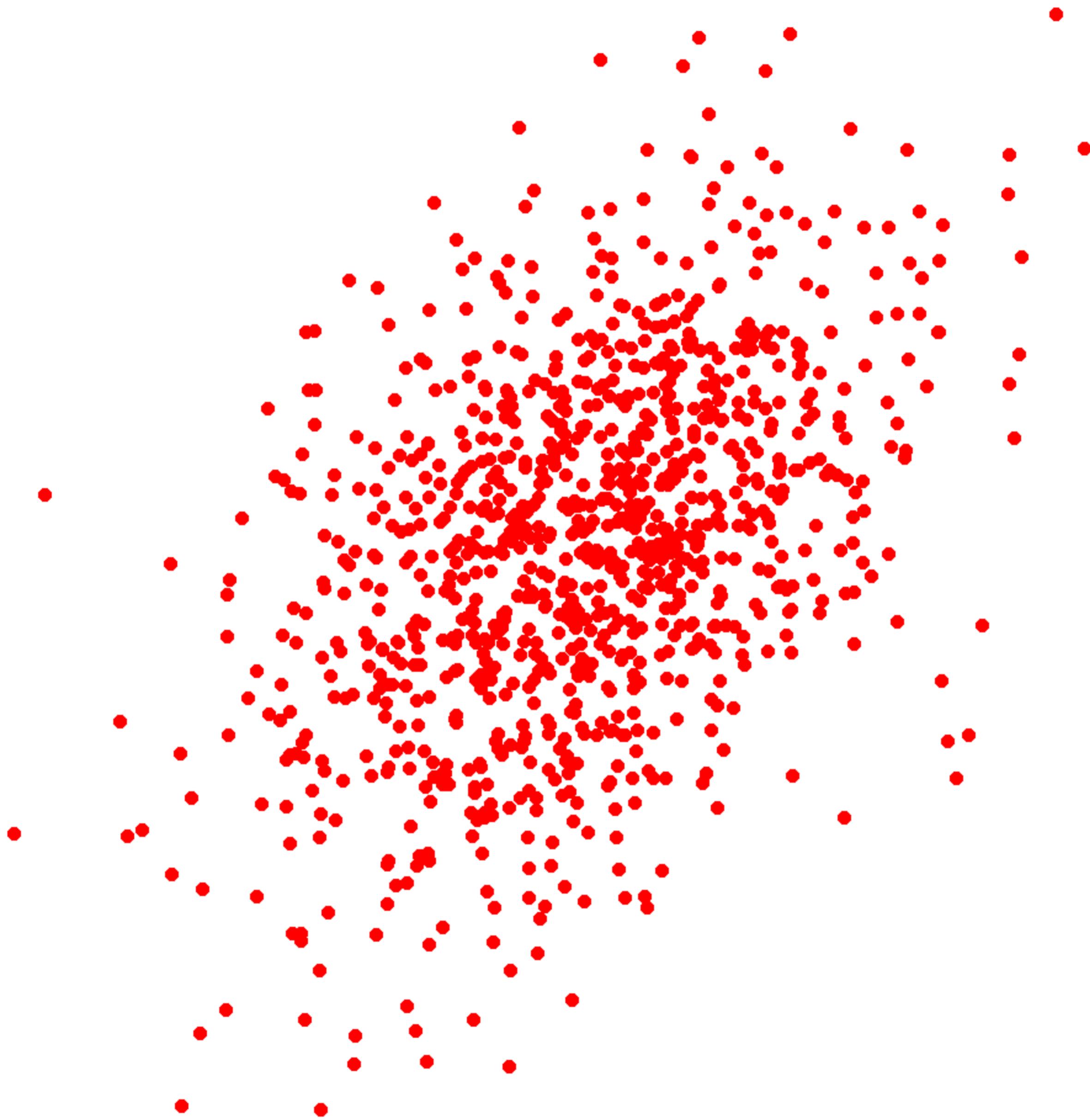
0.36



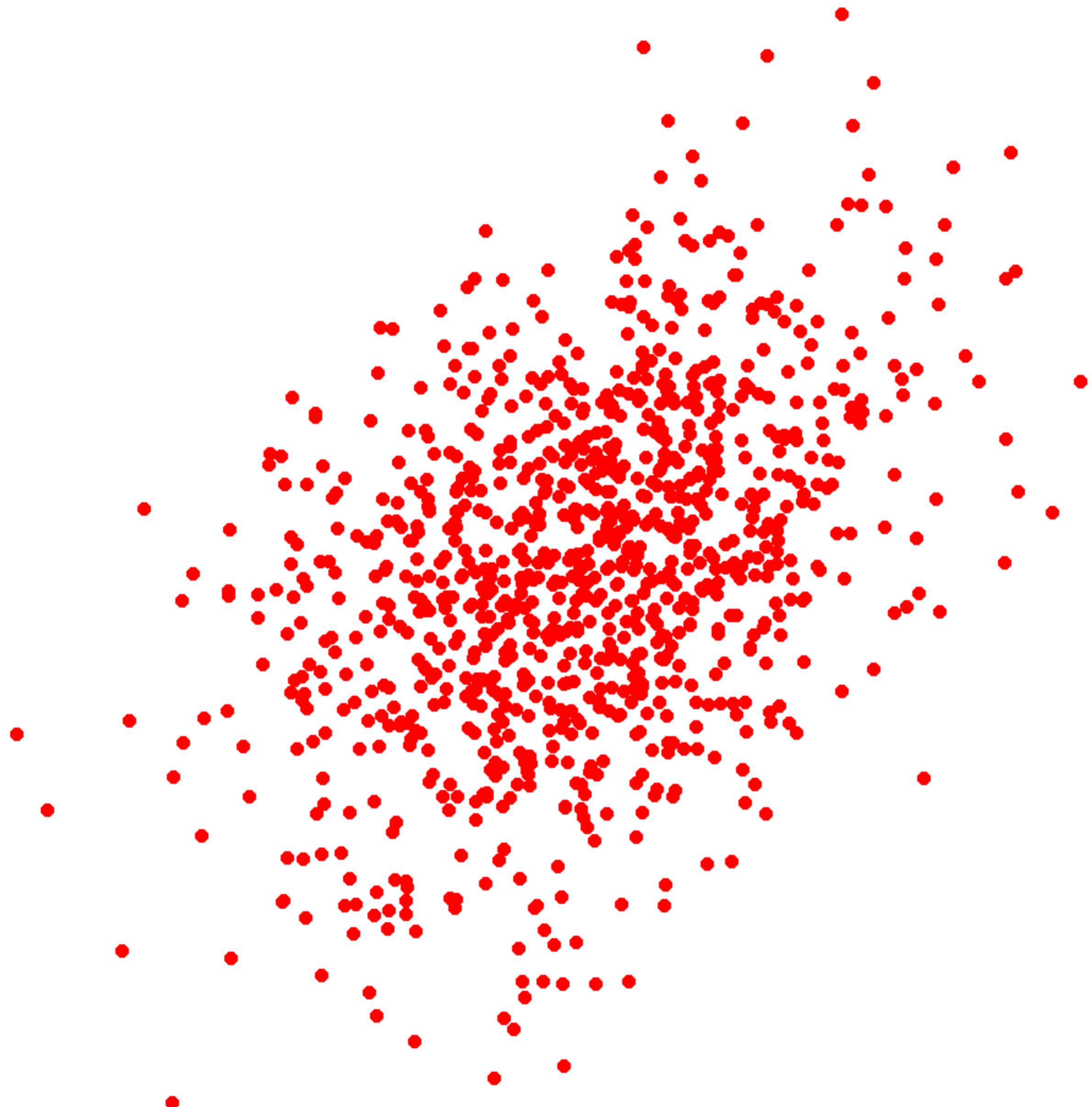
0.32



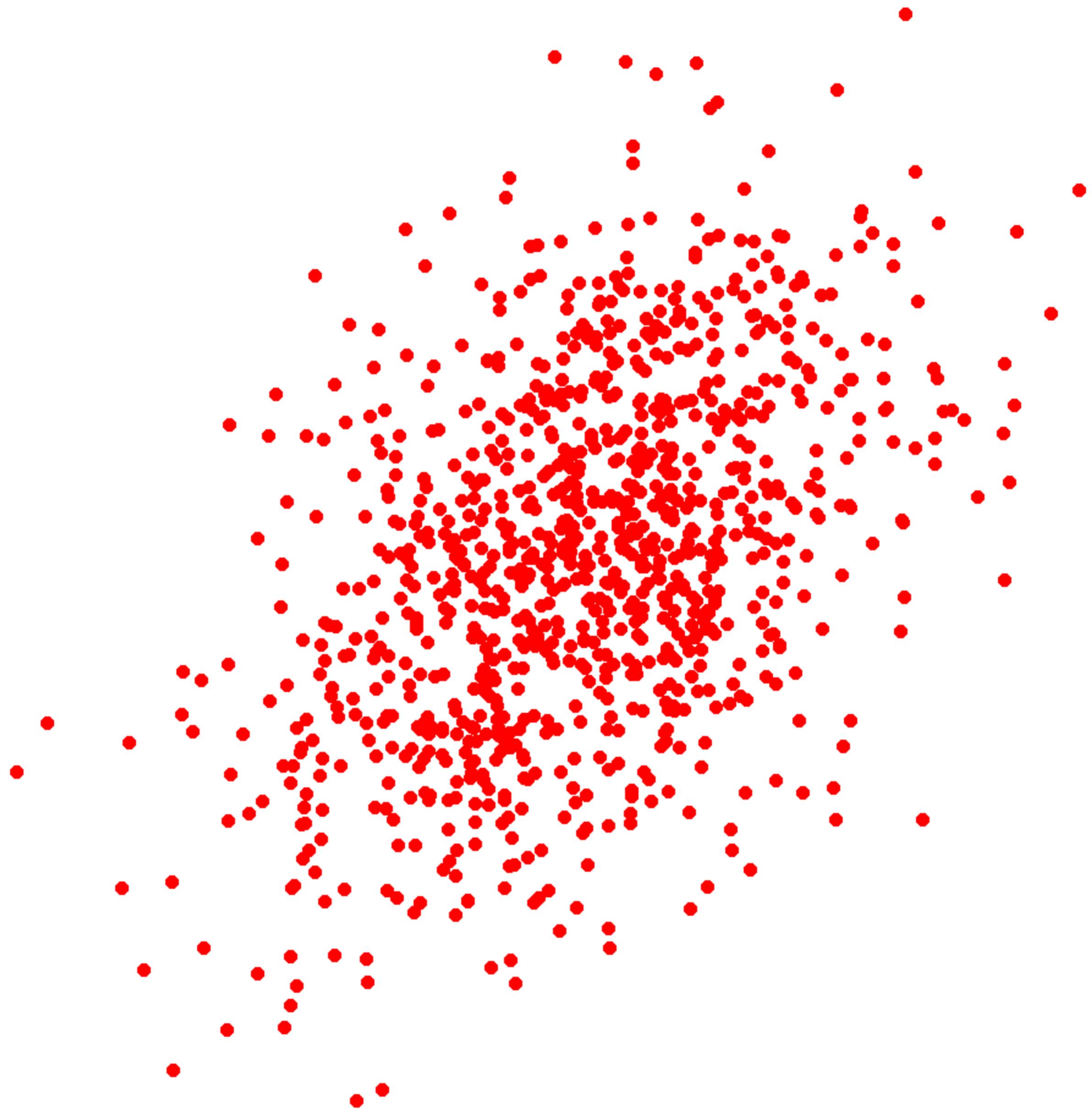
0.42



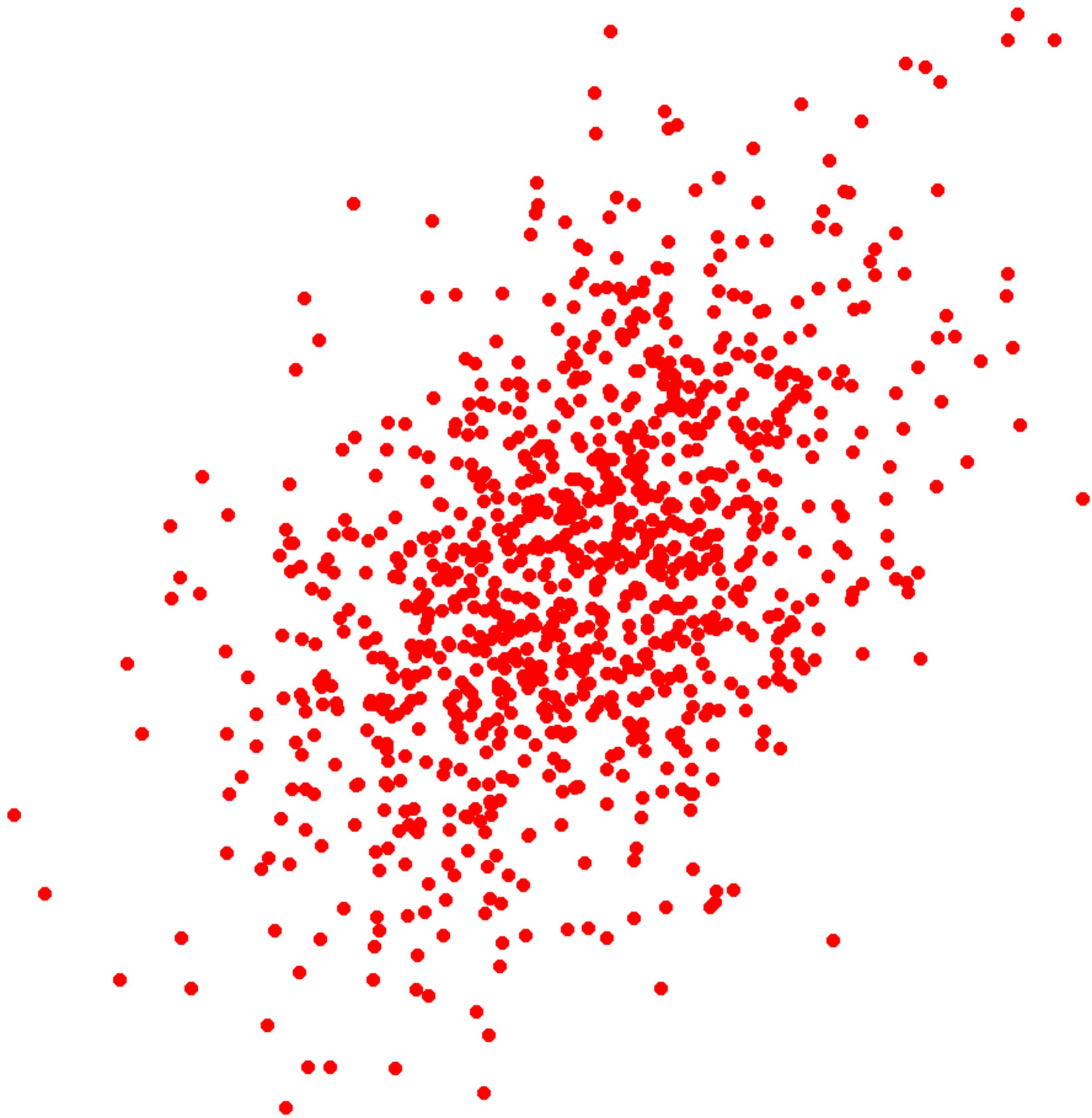
0.4



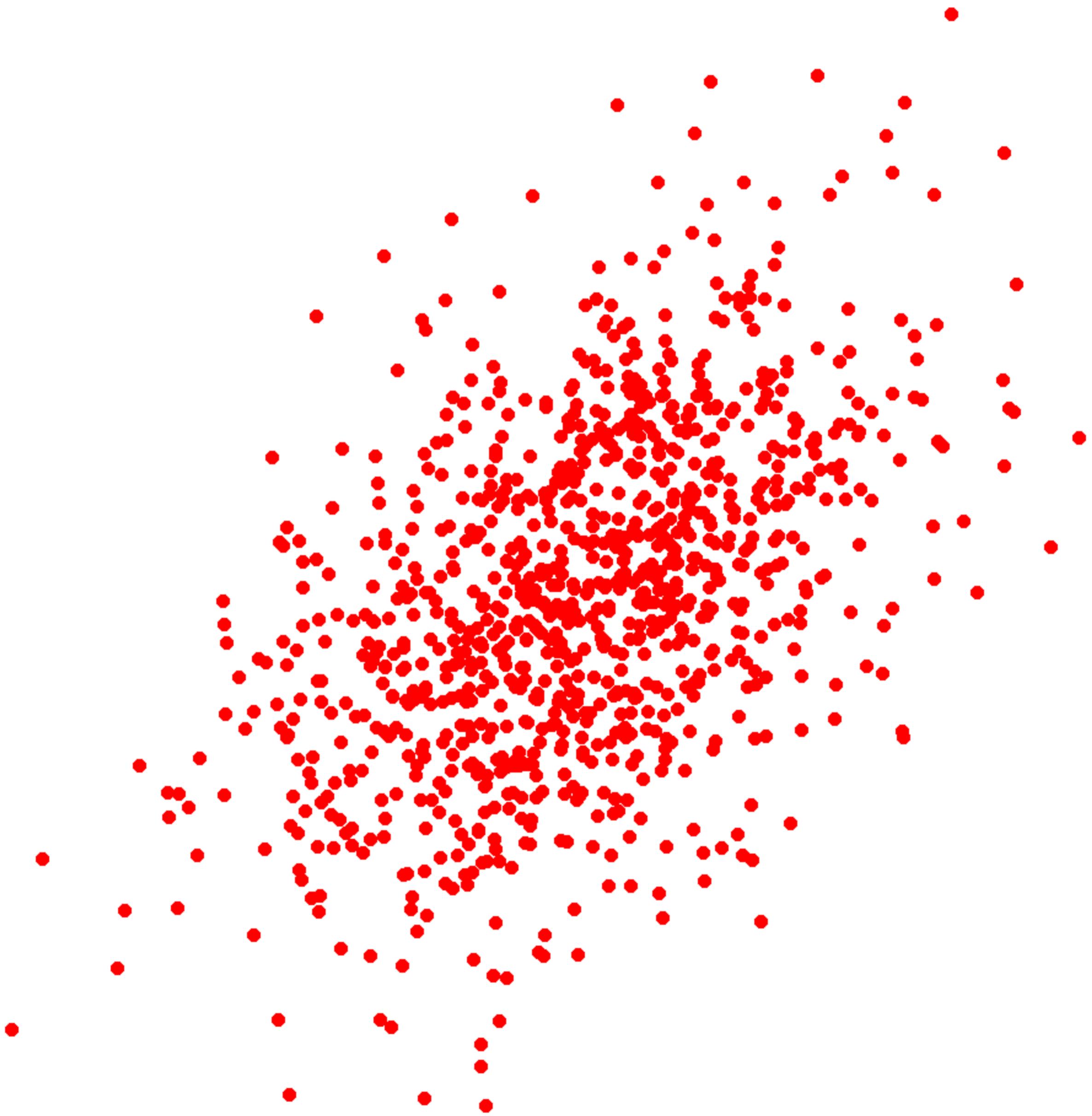
0.43



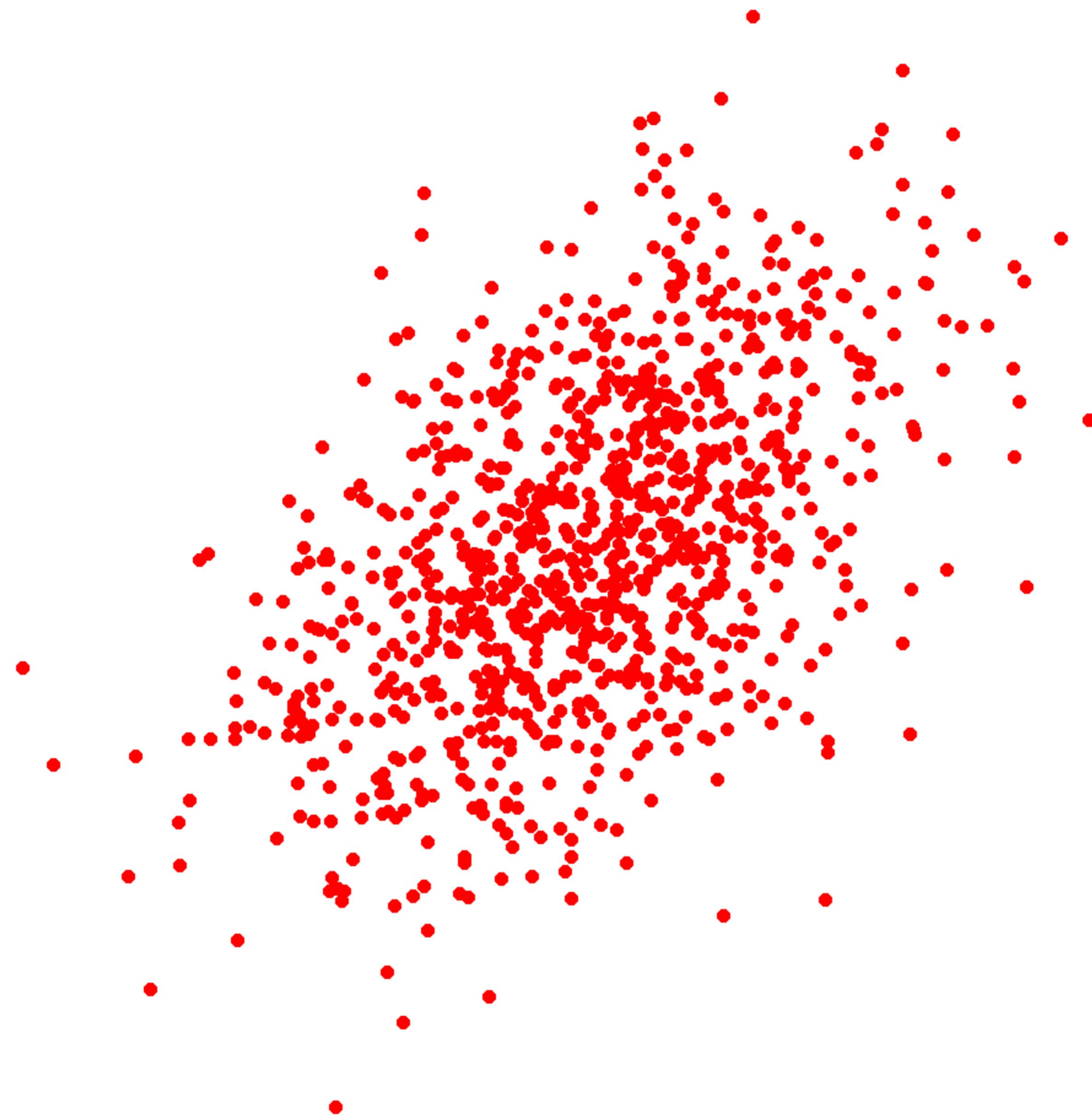
0.43



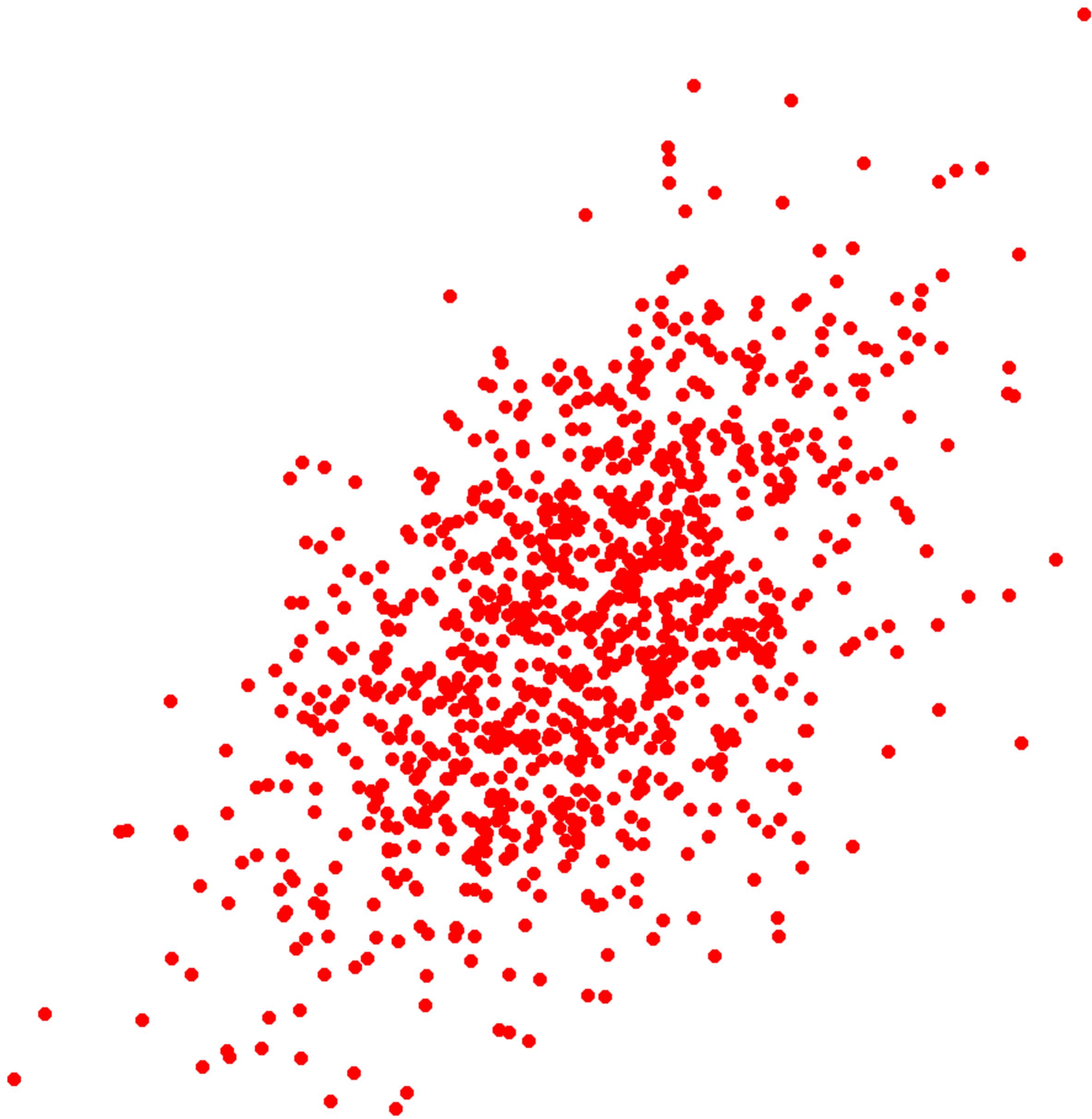
0.48



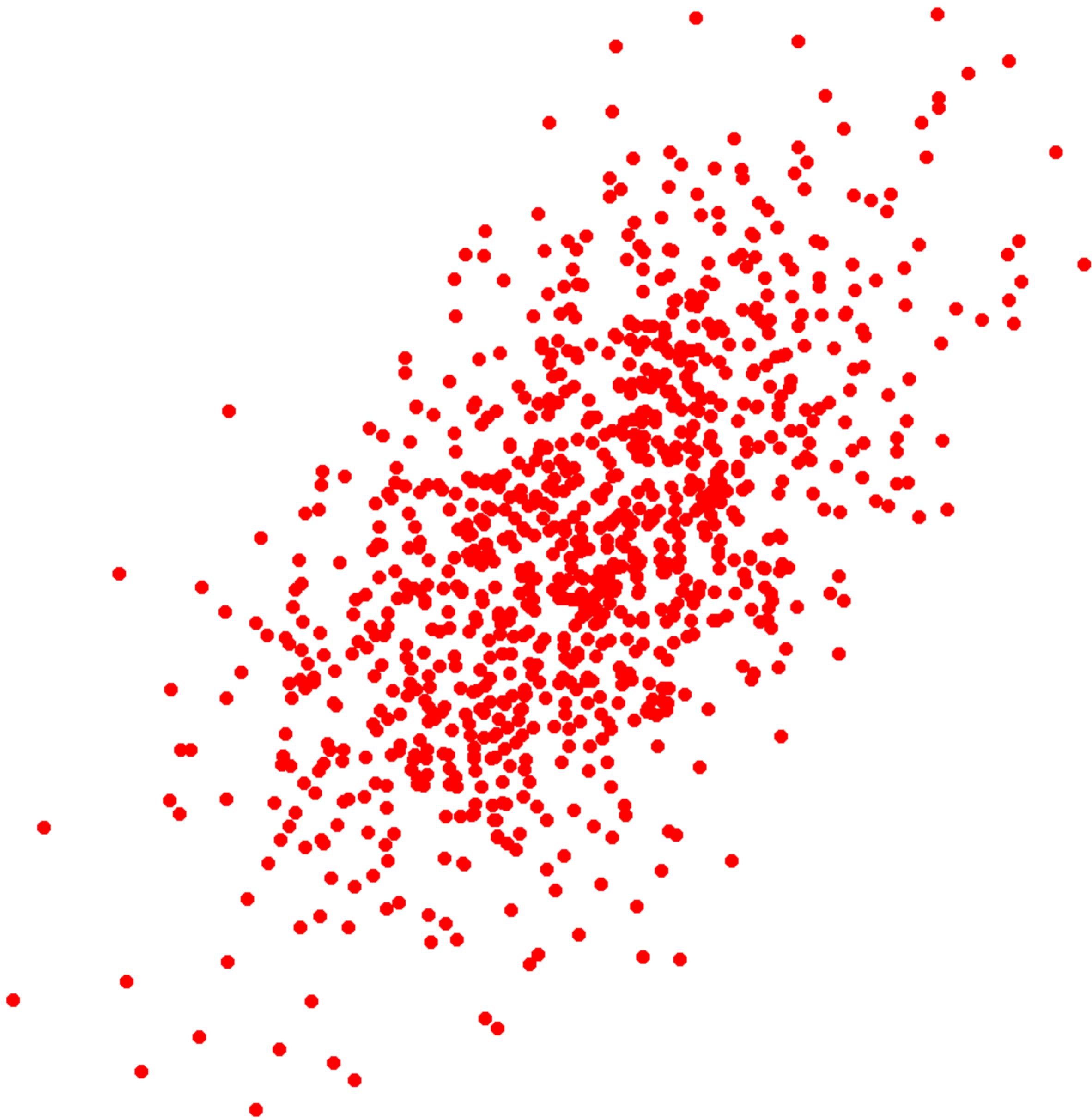
0.5



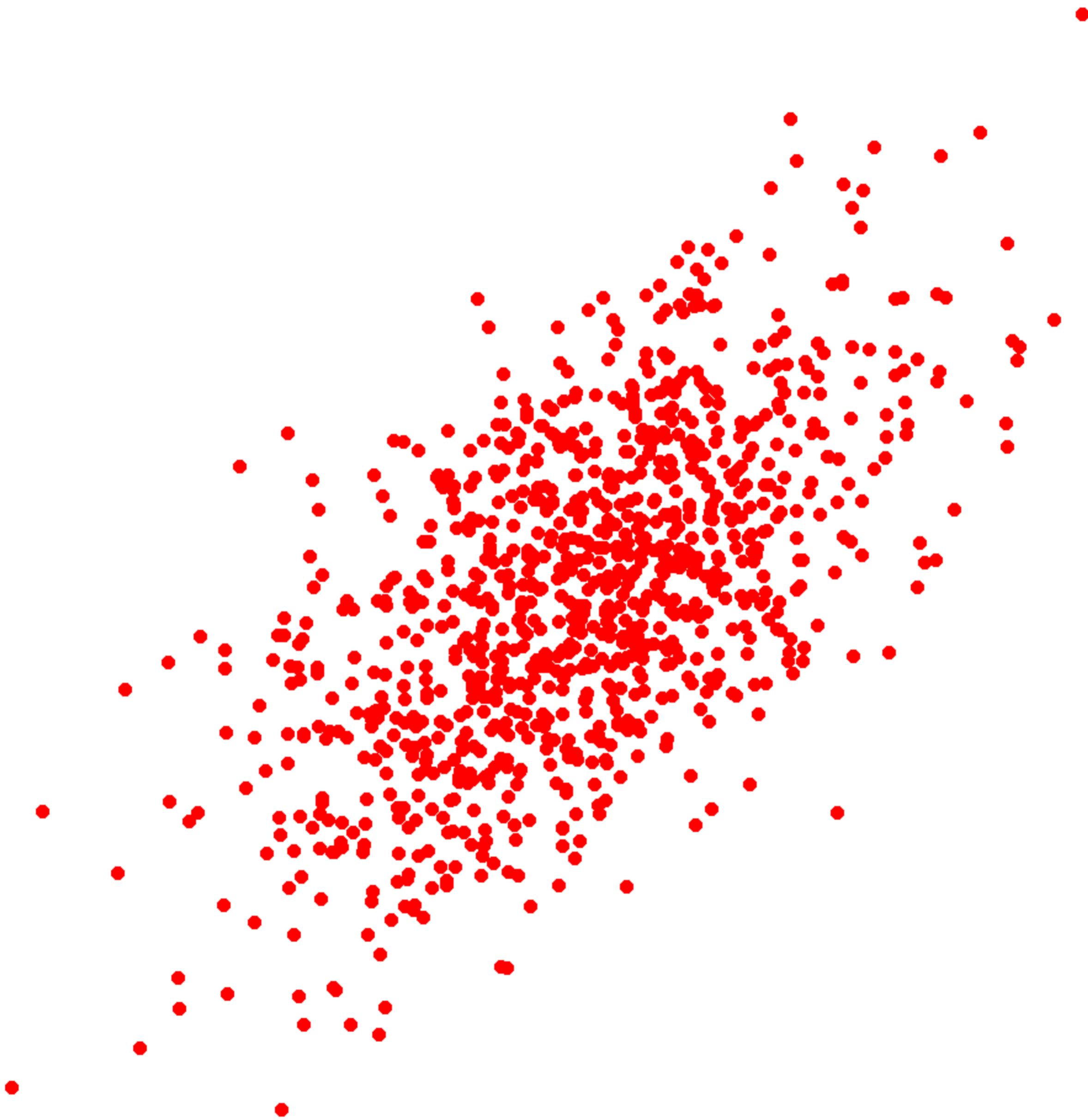
0.5



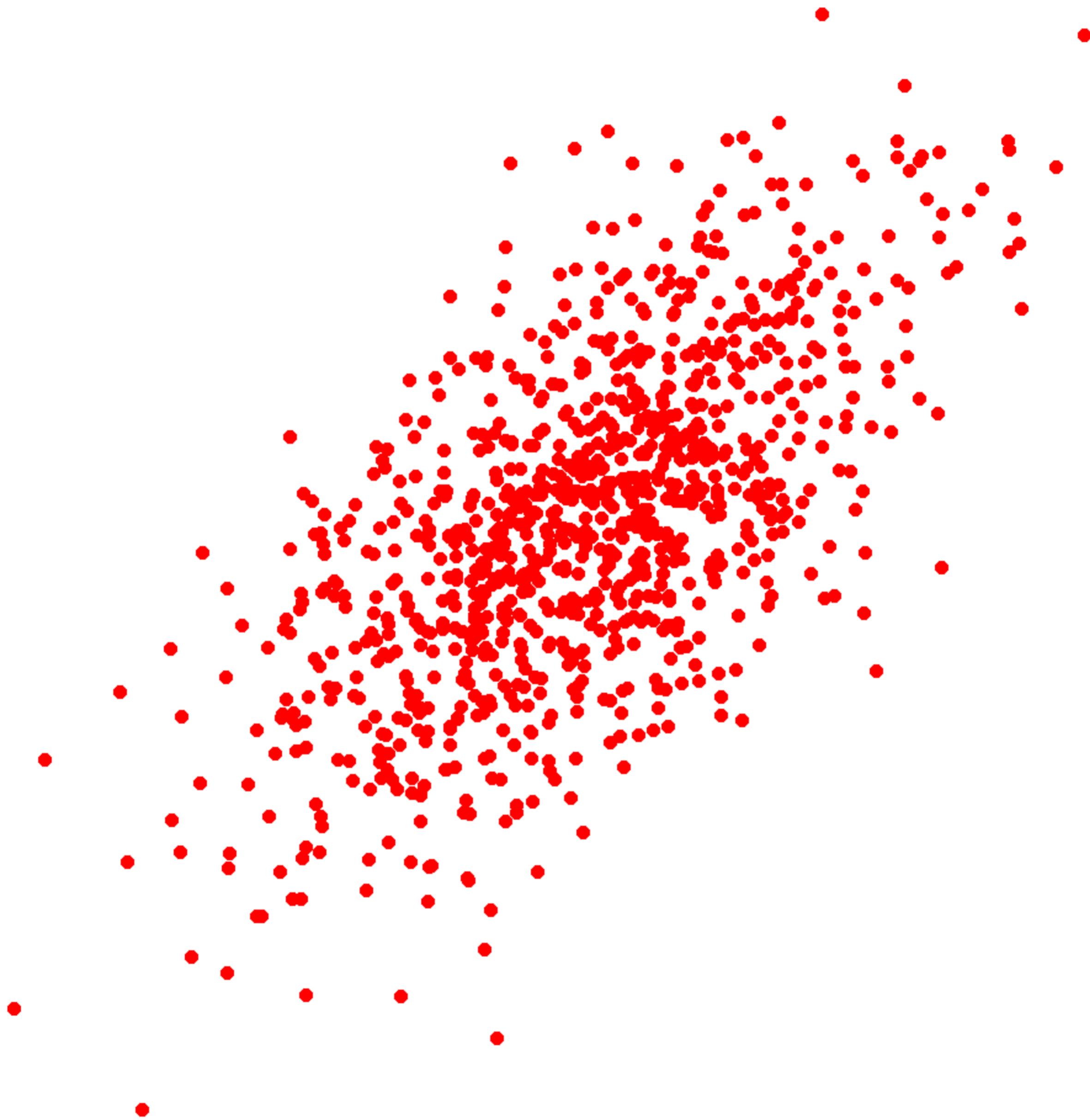
0.57



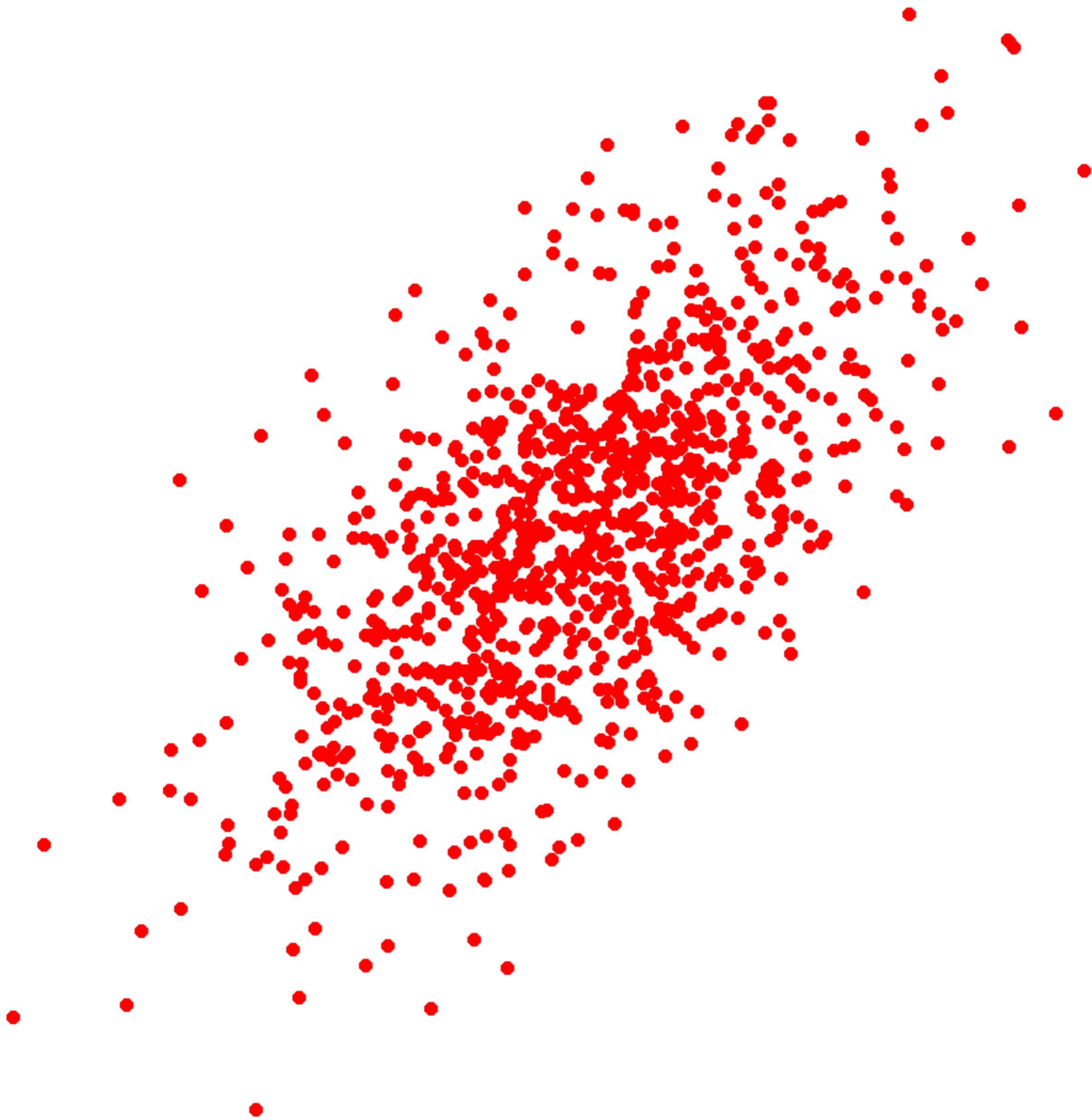
0.58



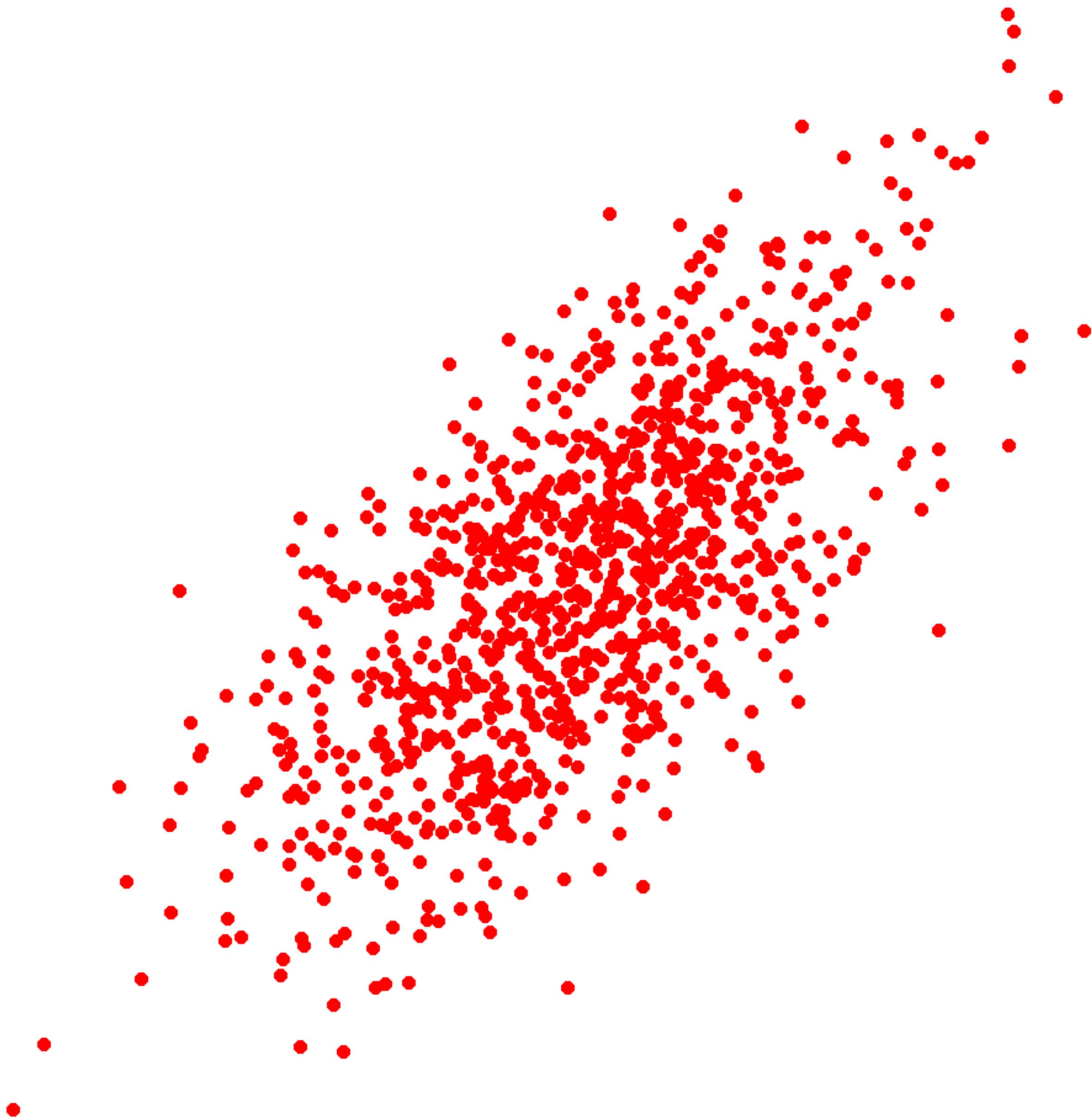
0.61



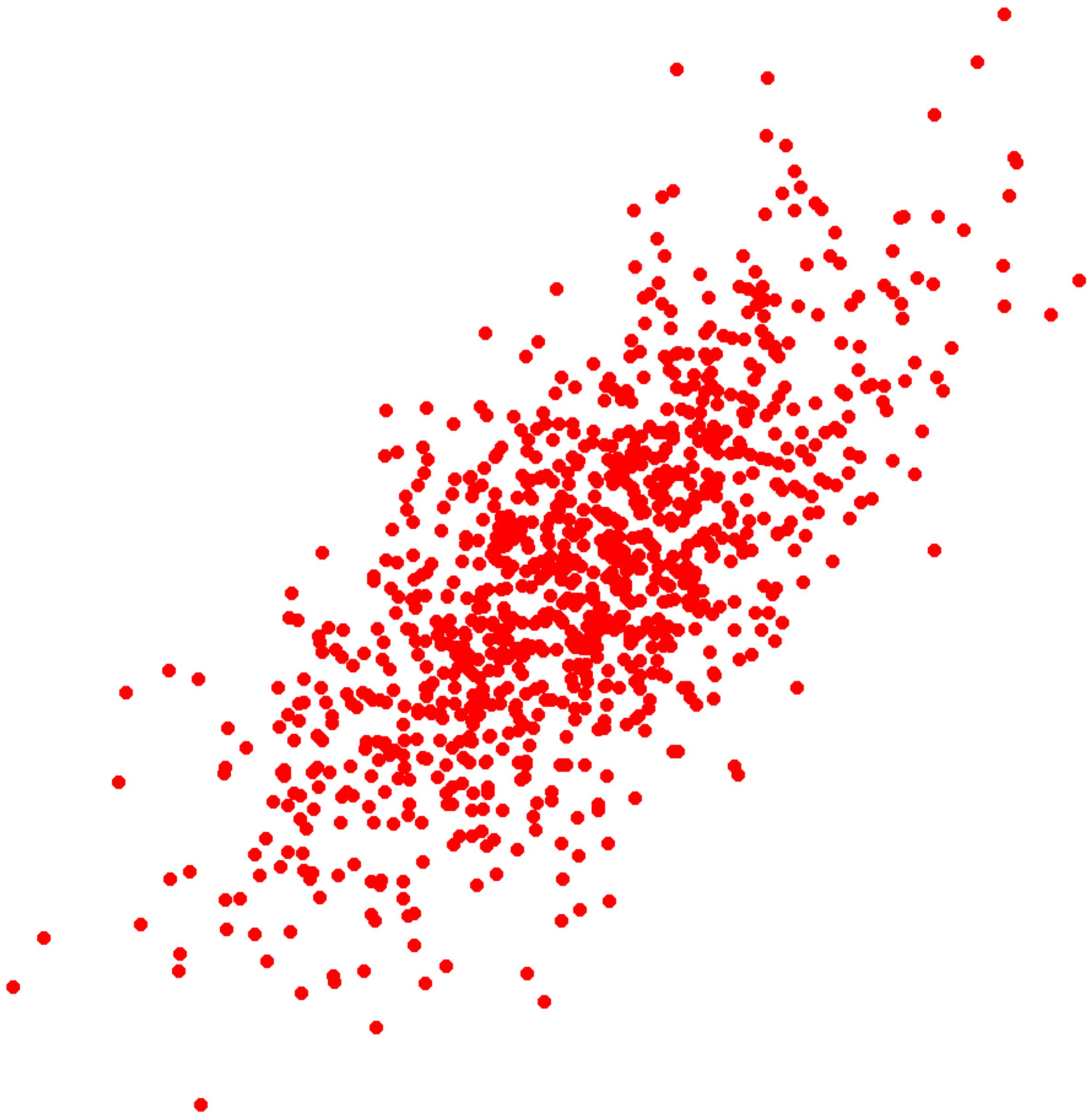
0.62



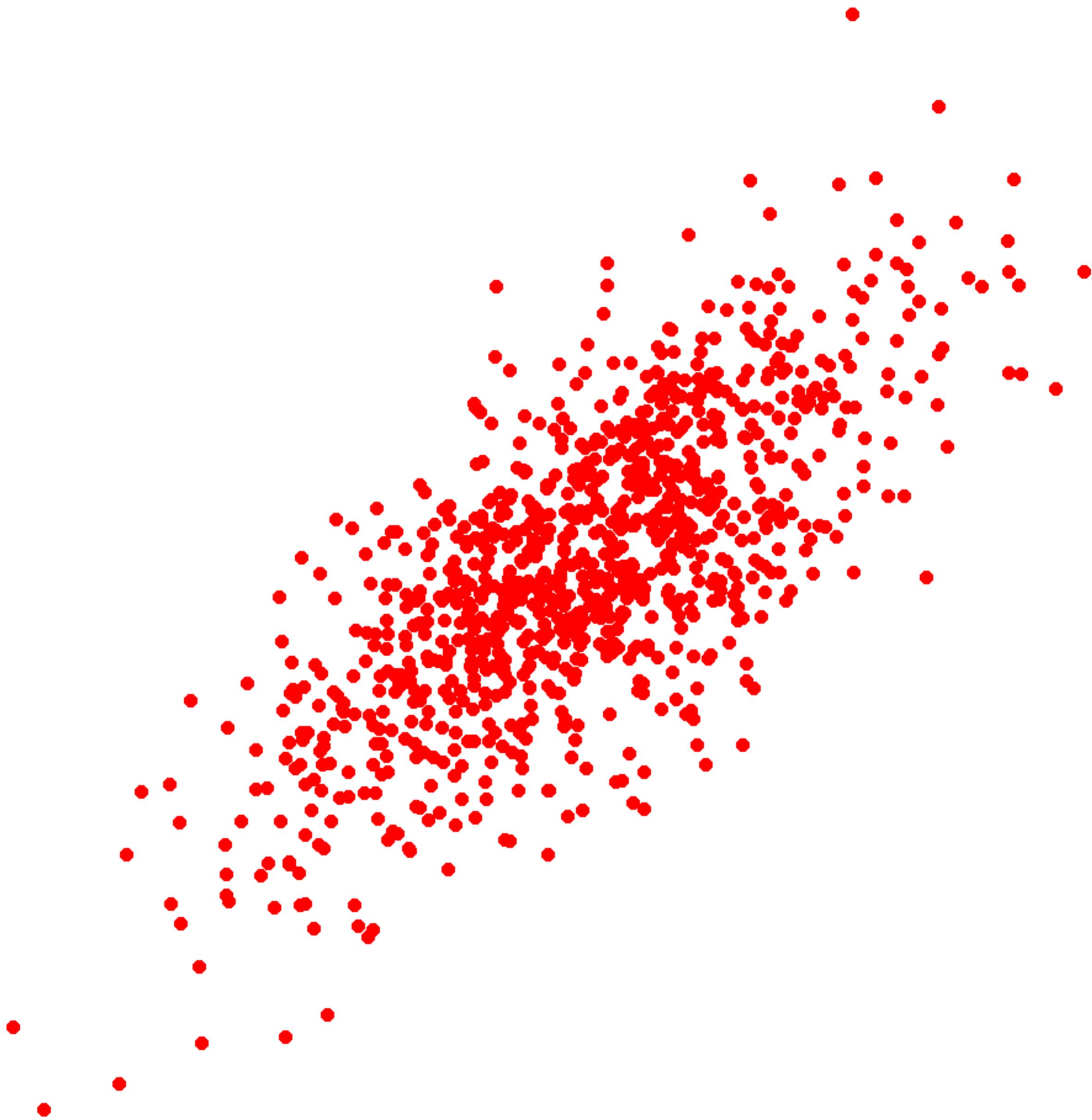
0.66



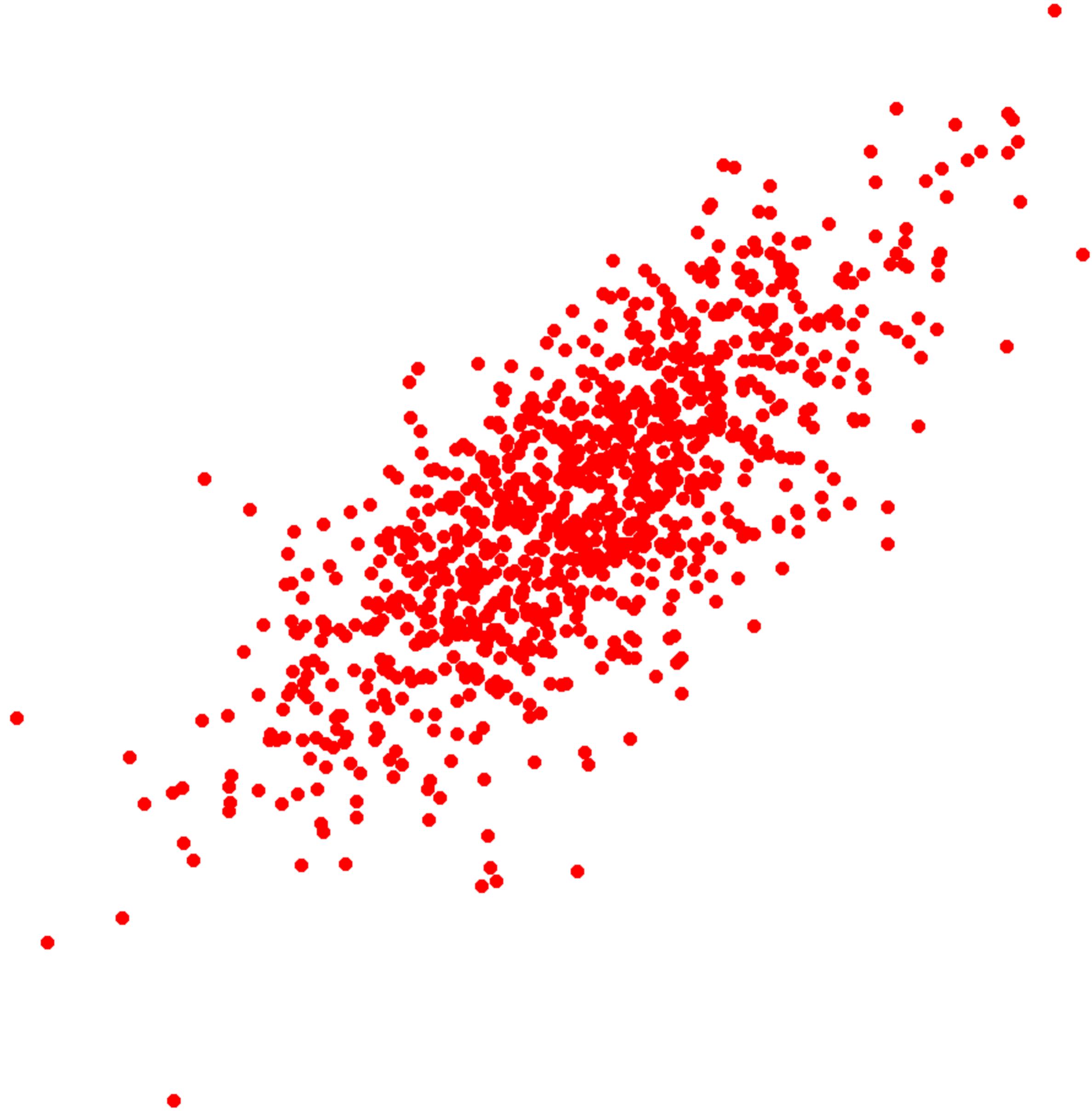
0.67



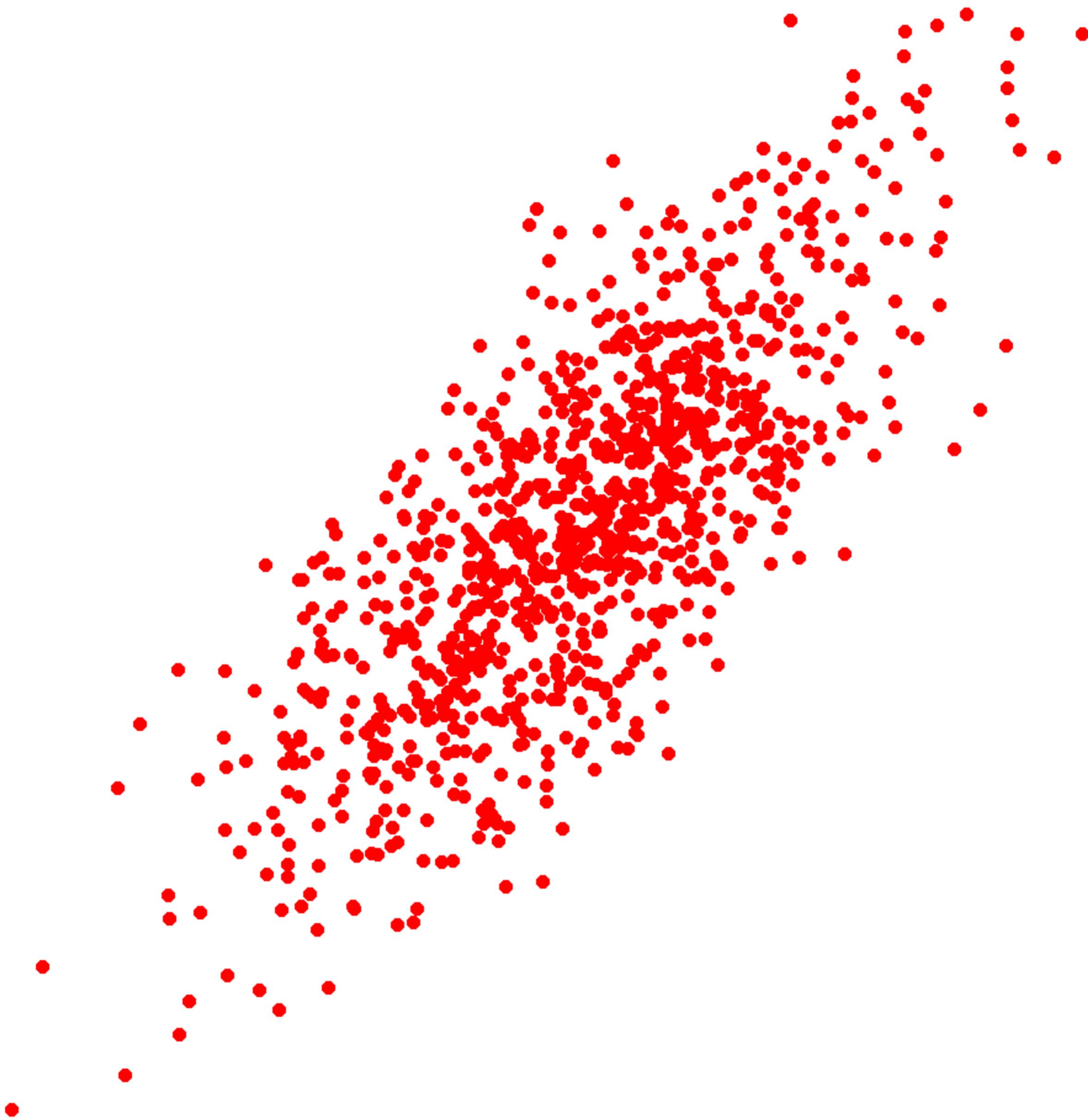
0.69



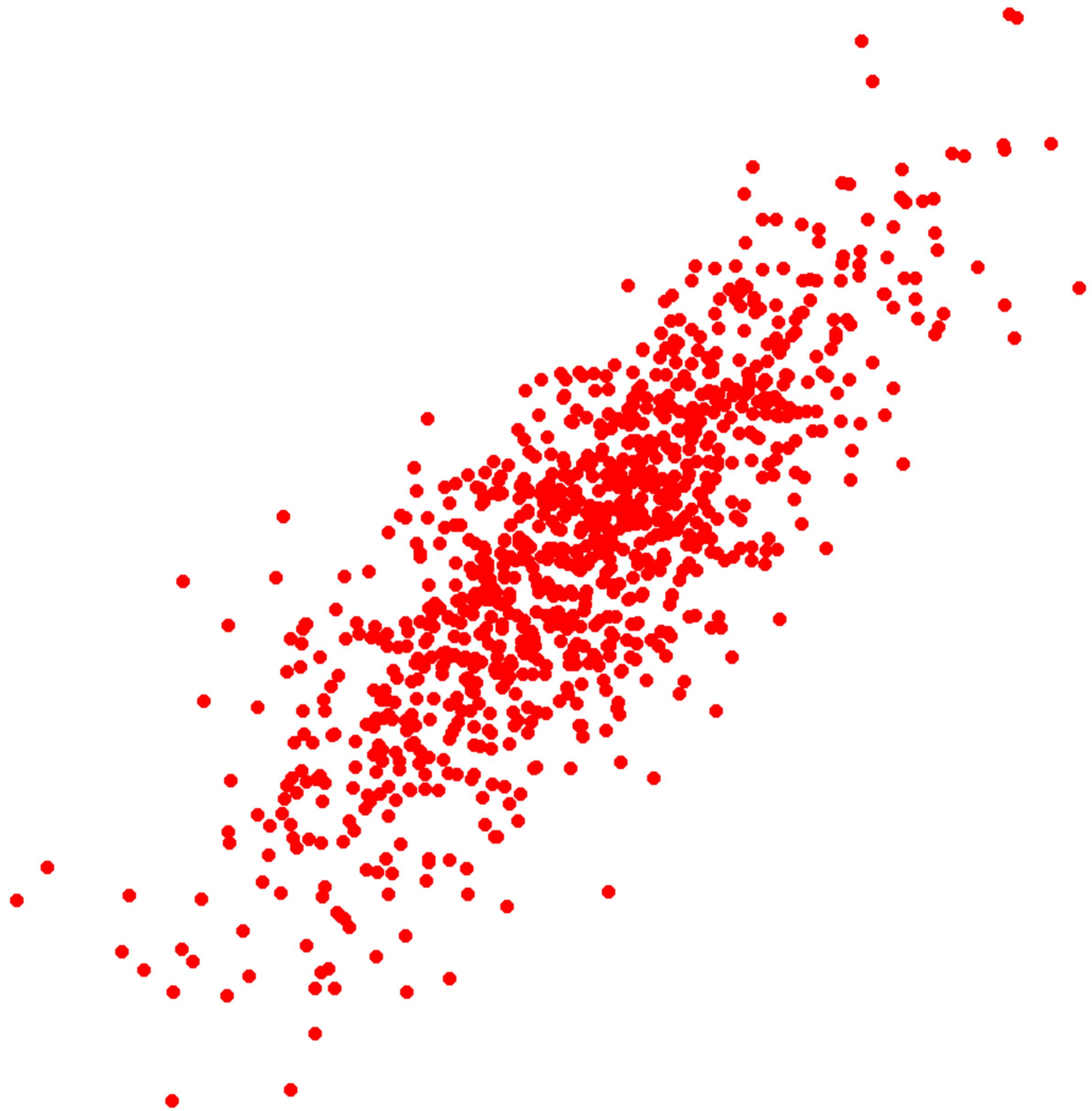
0.74



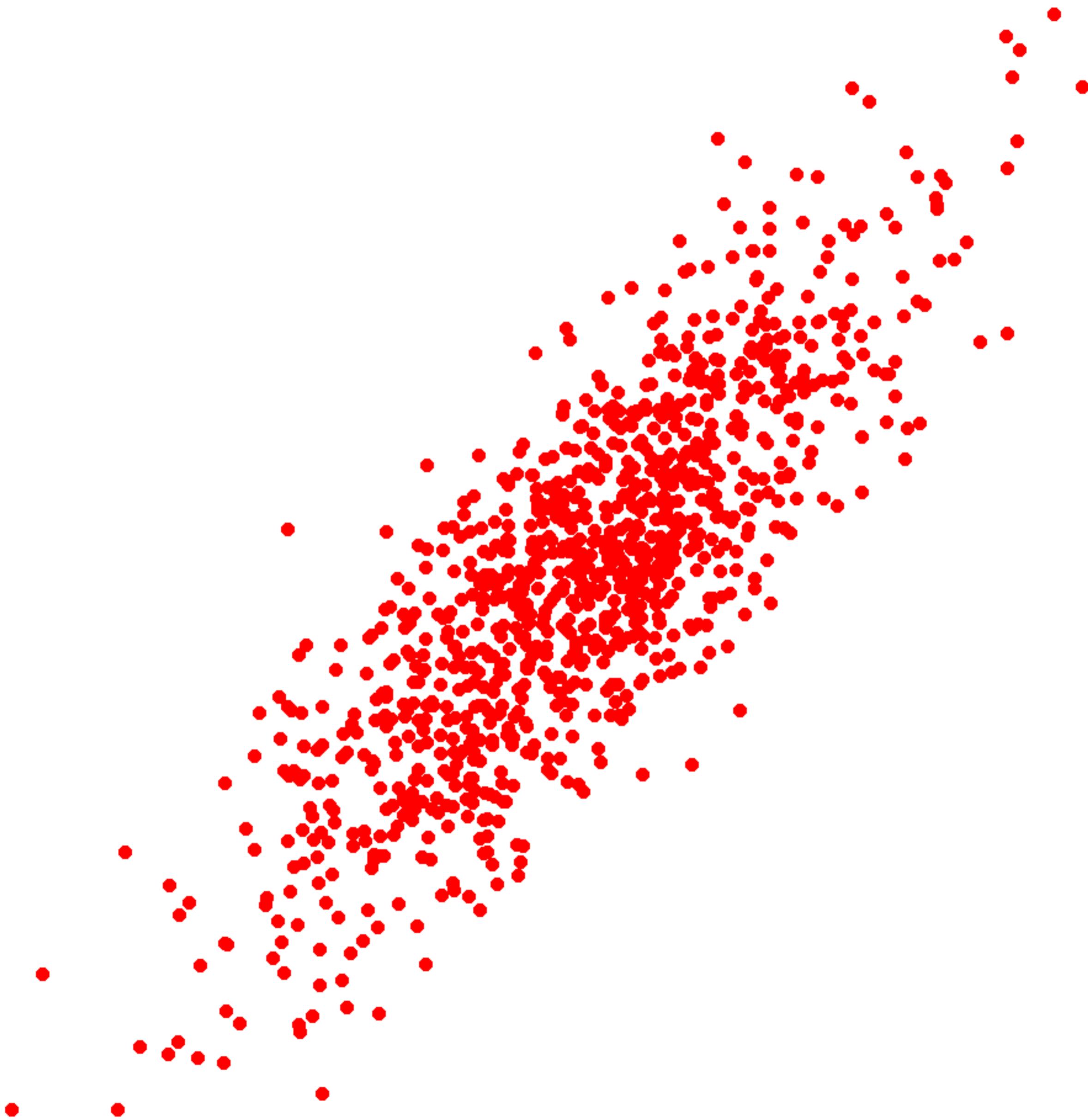
0.75



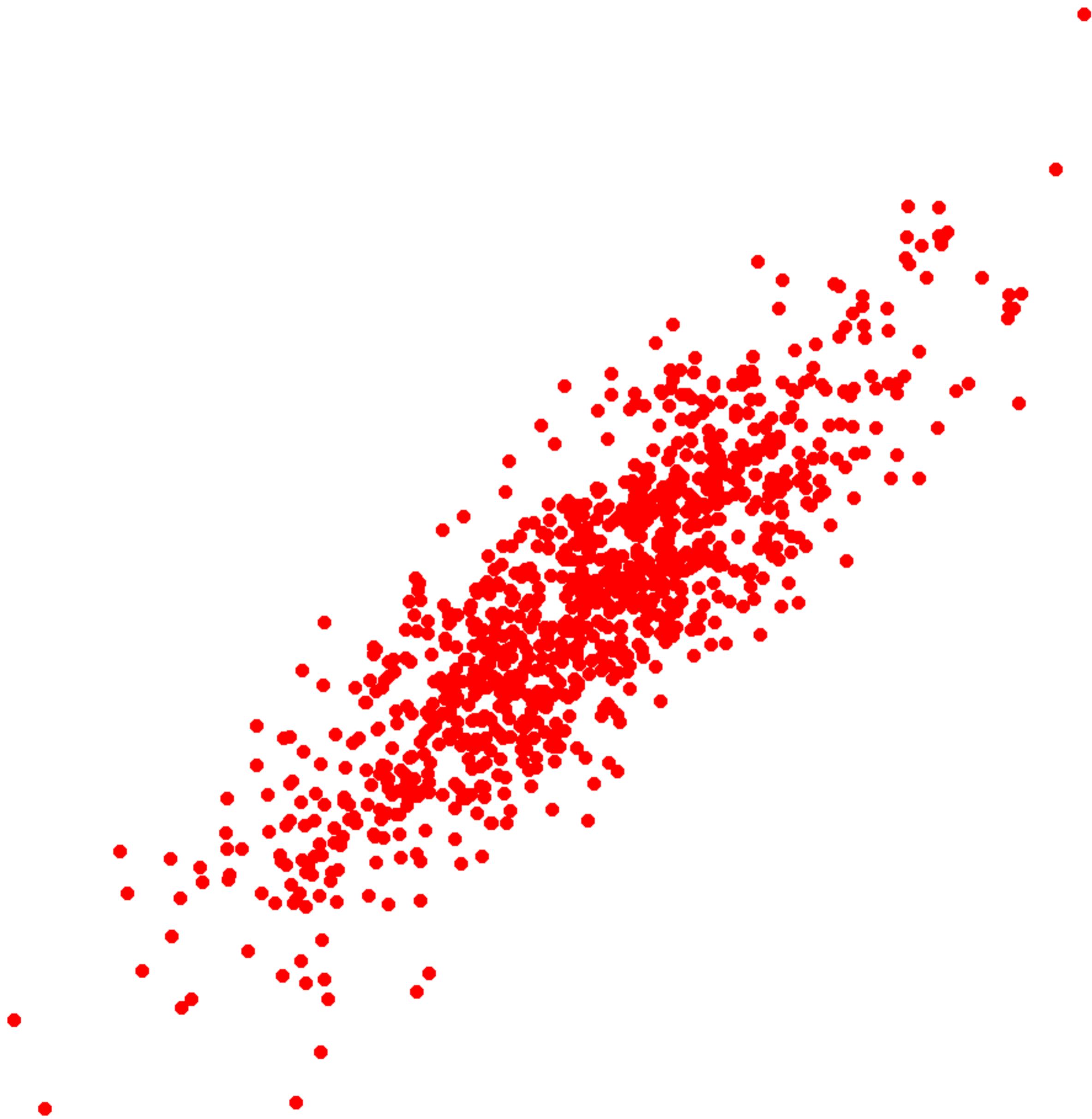
0.79



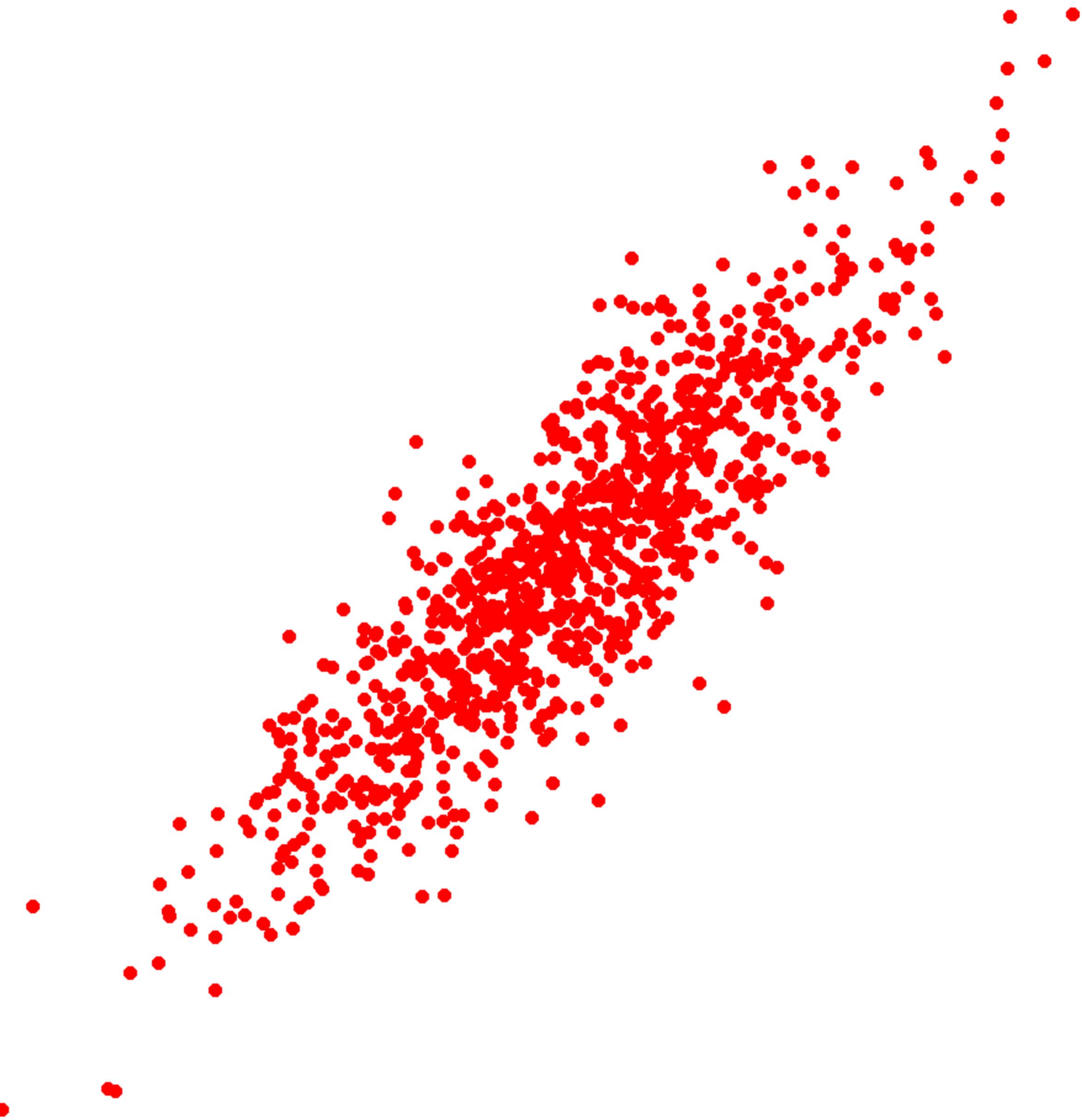
0.81



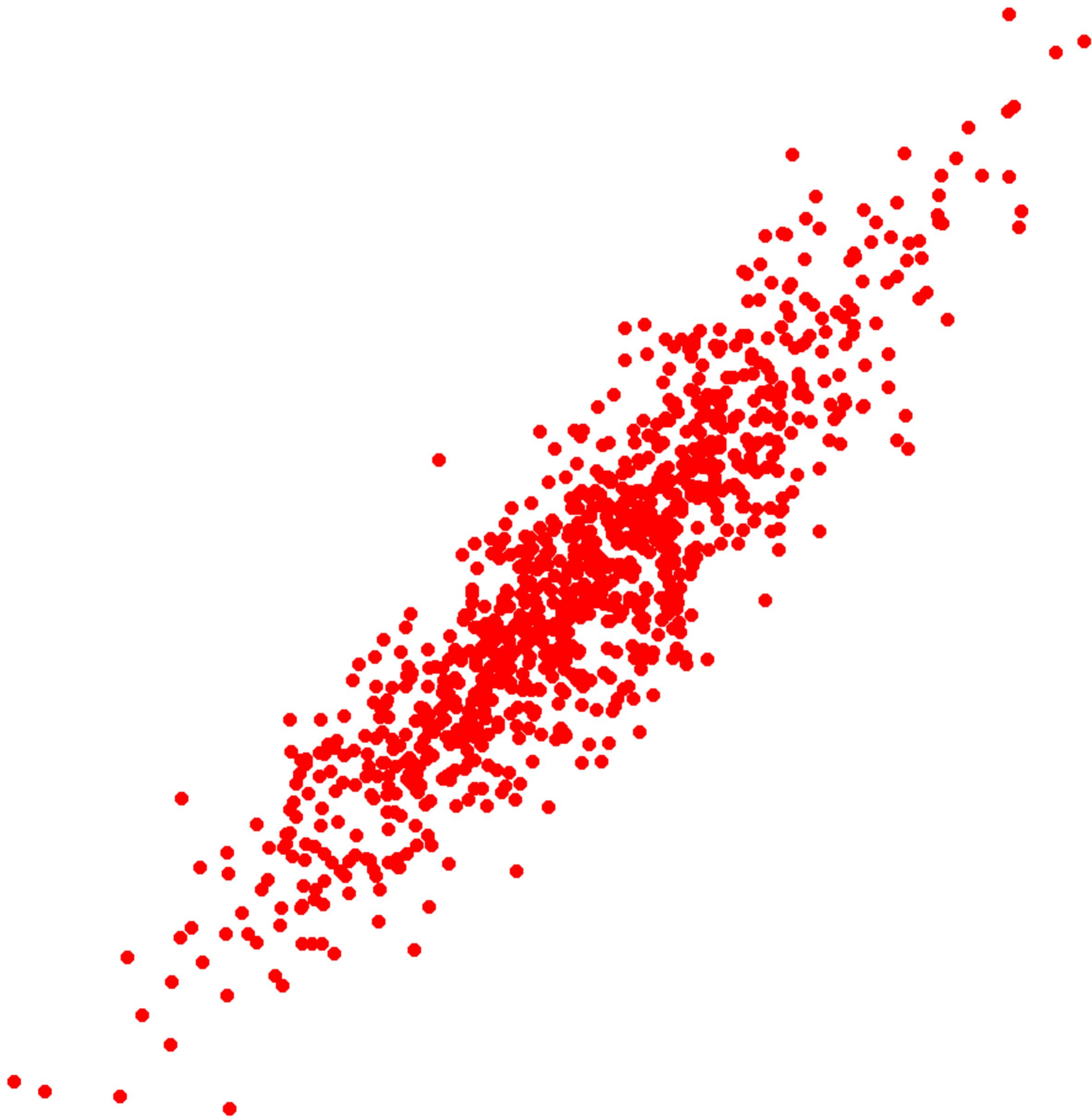
0.85



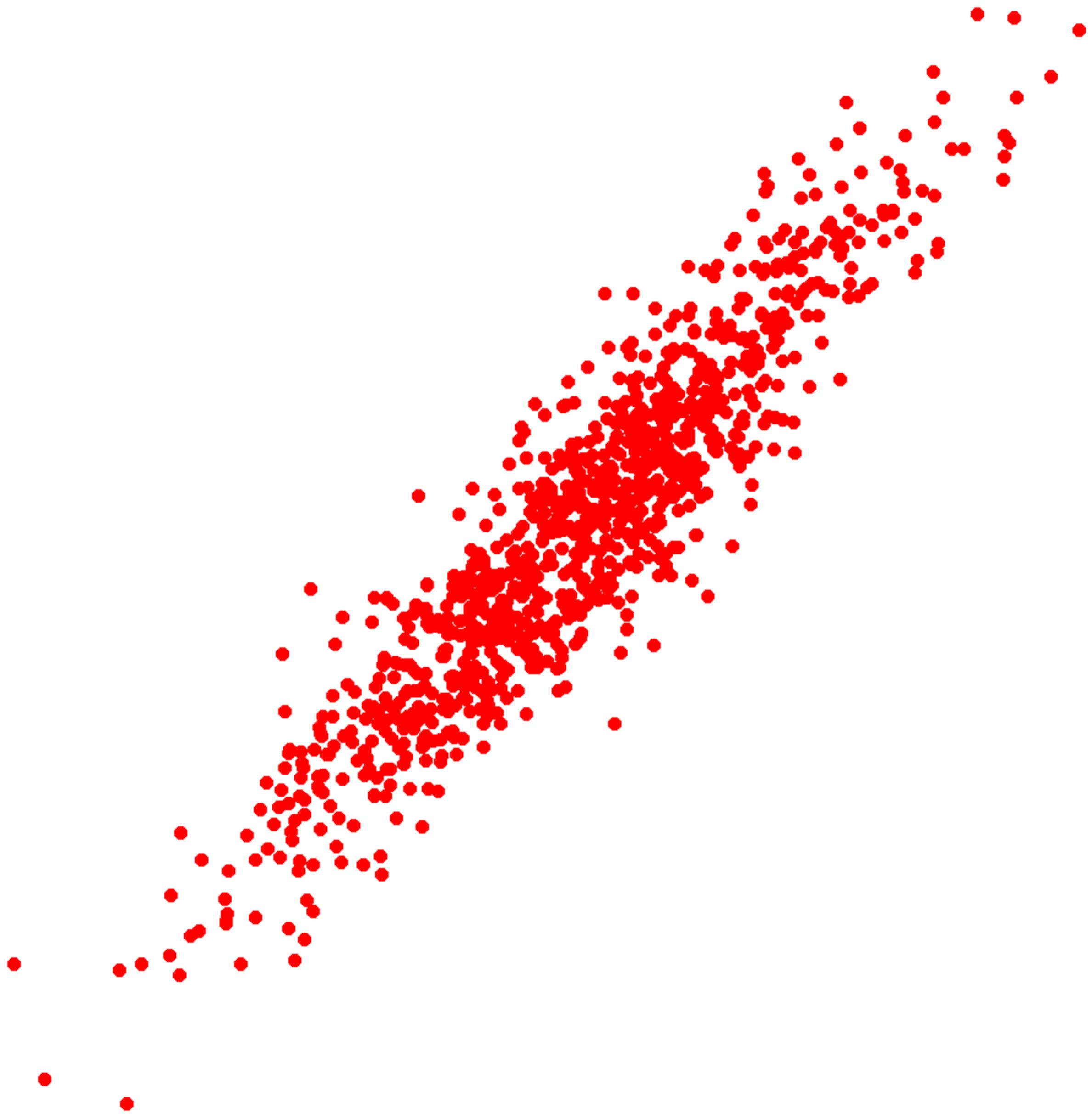
0.86



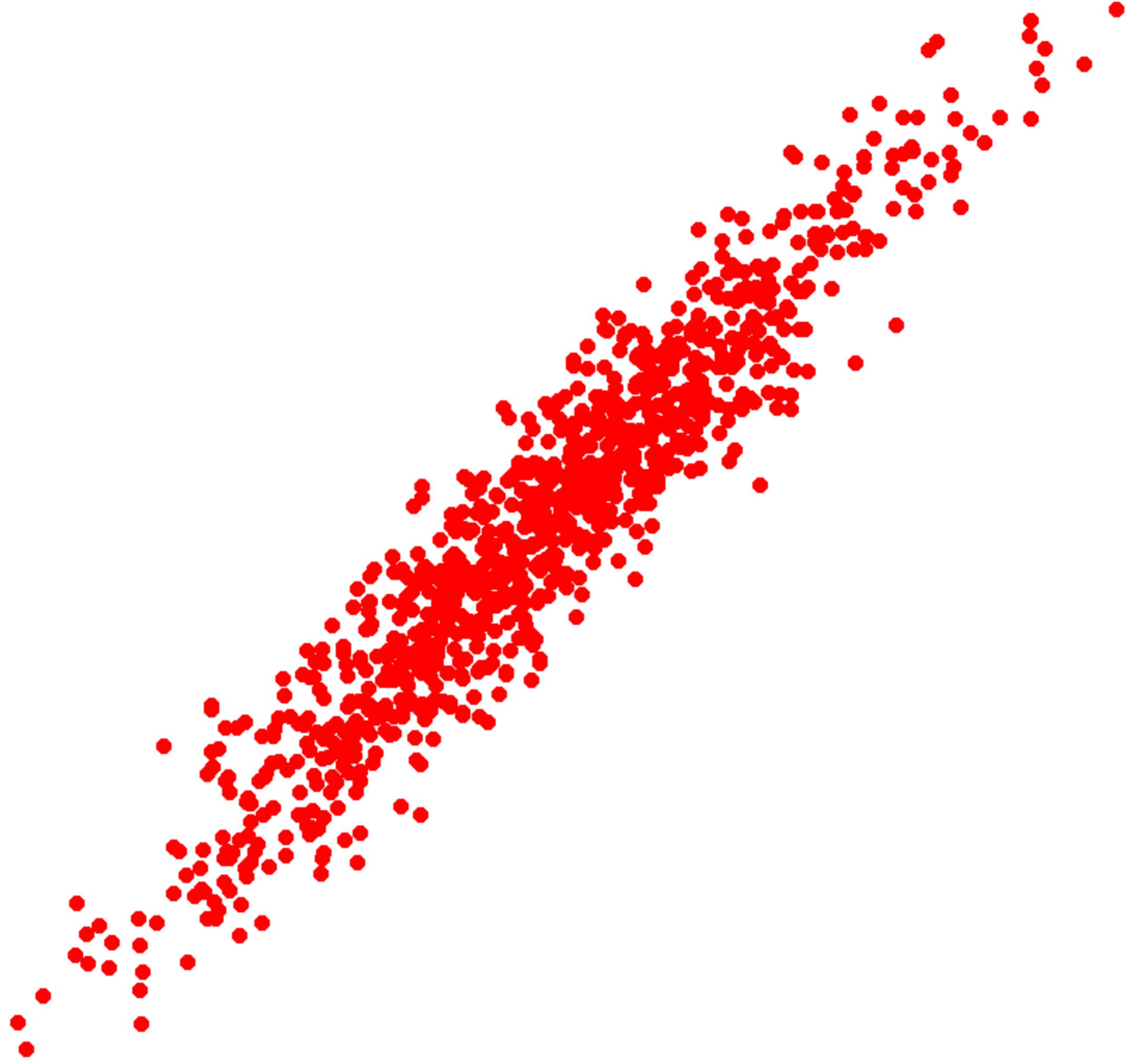
0.89



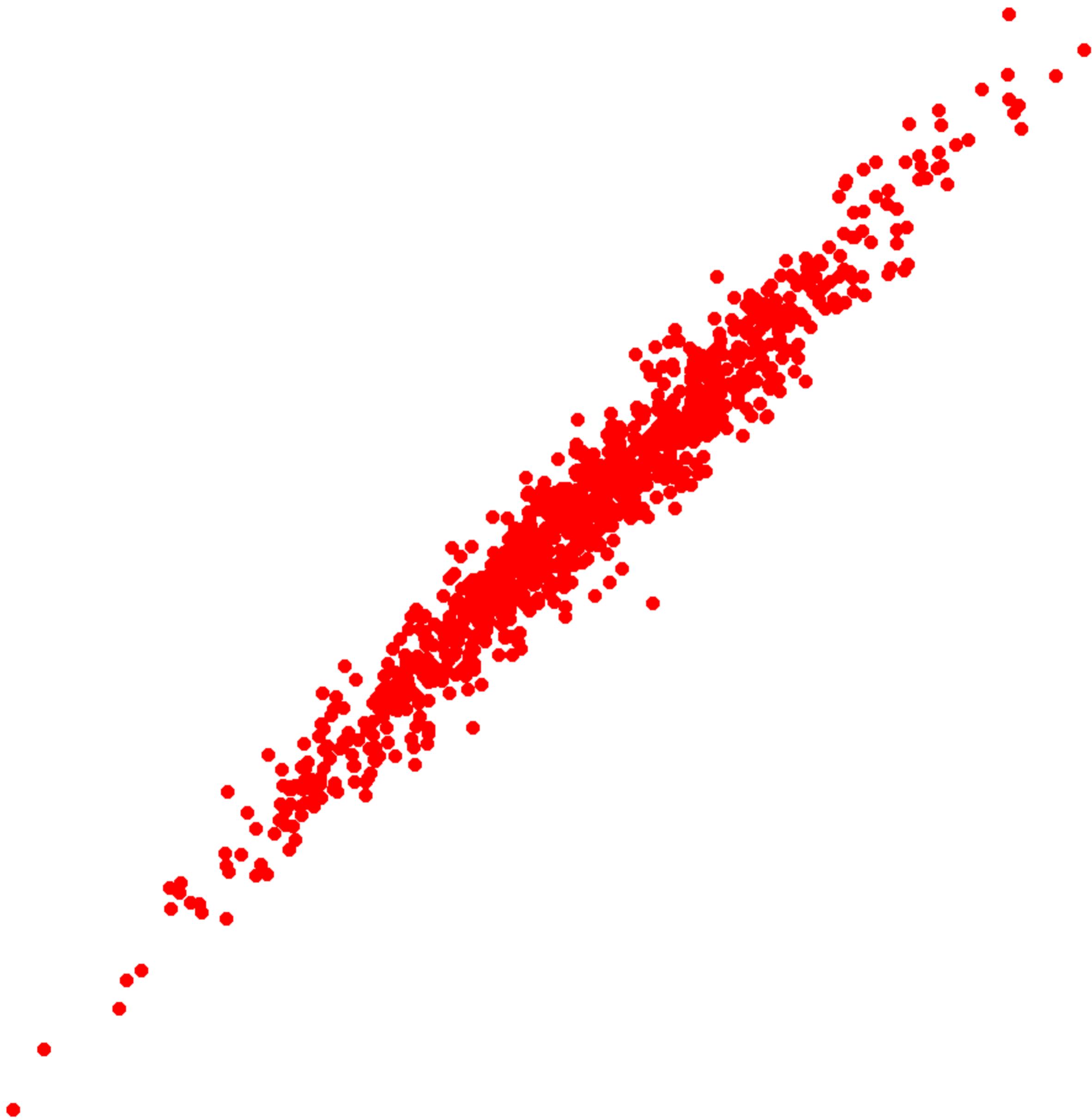
0.91



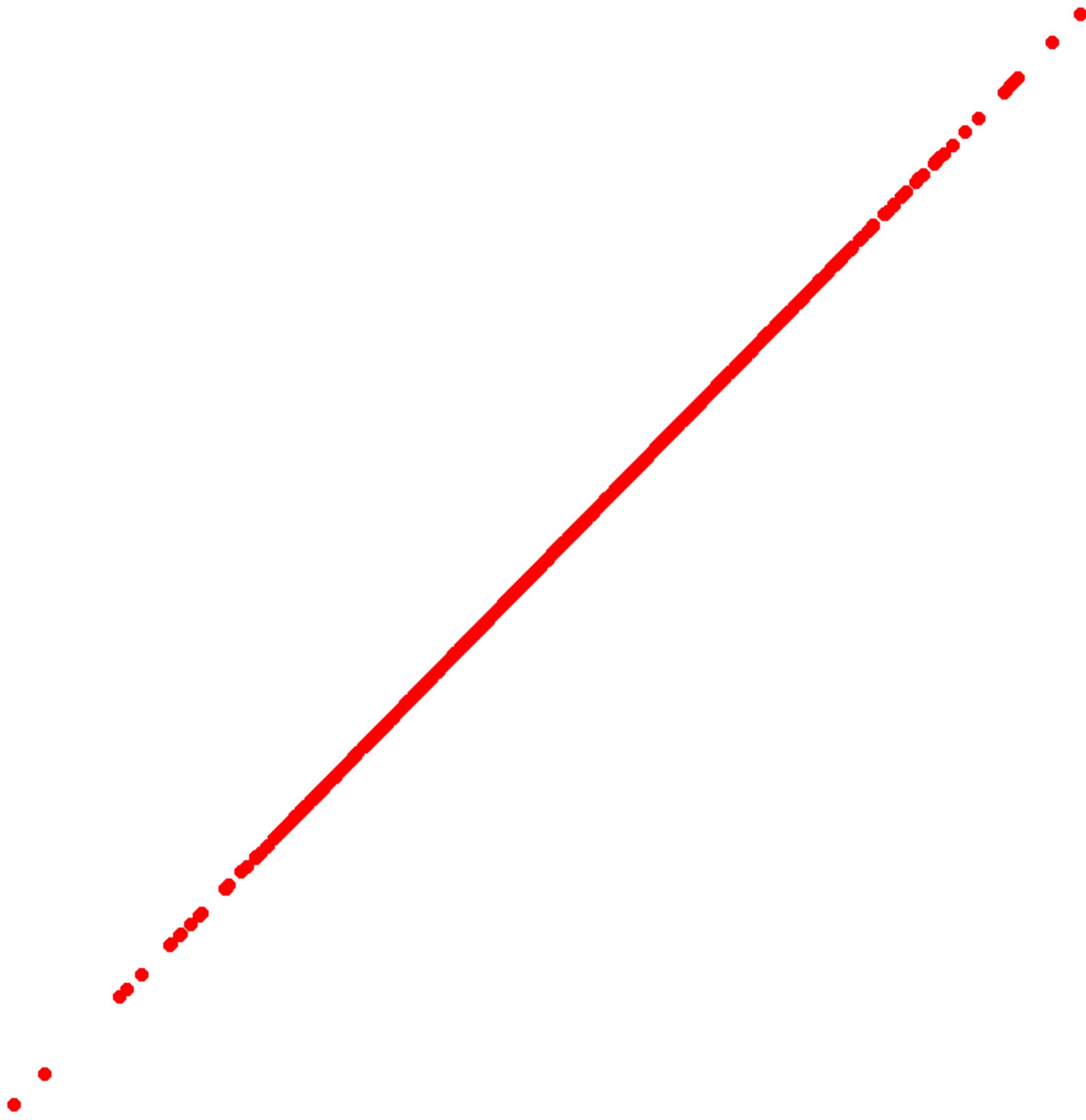
0.94



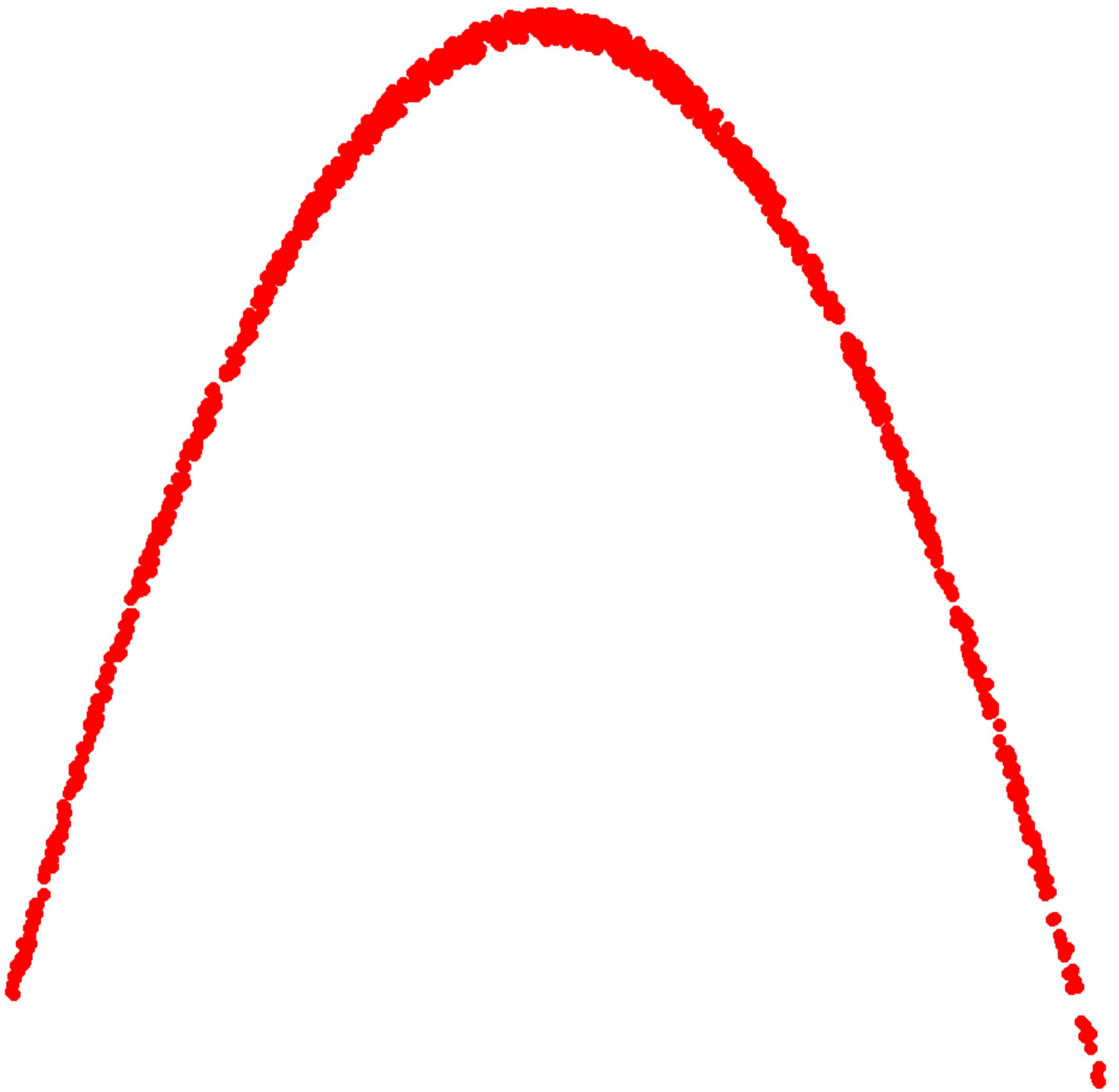
0.97



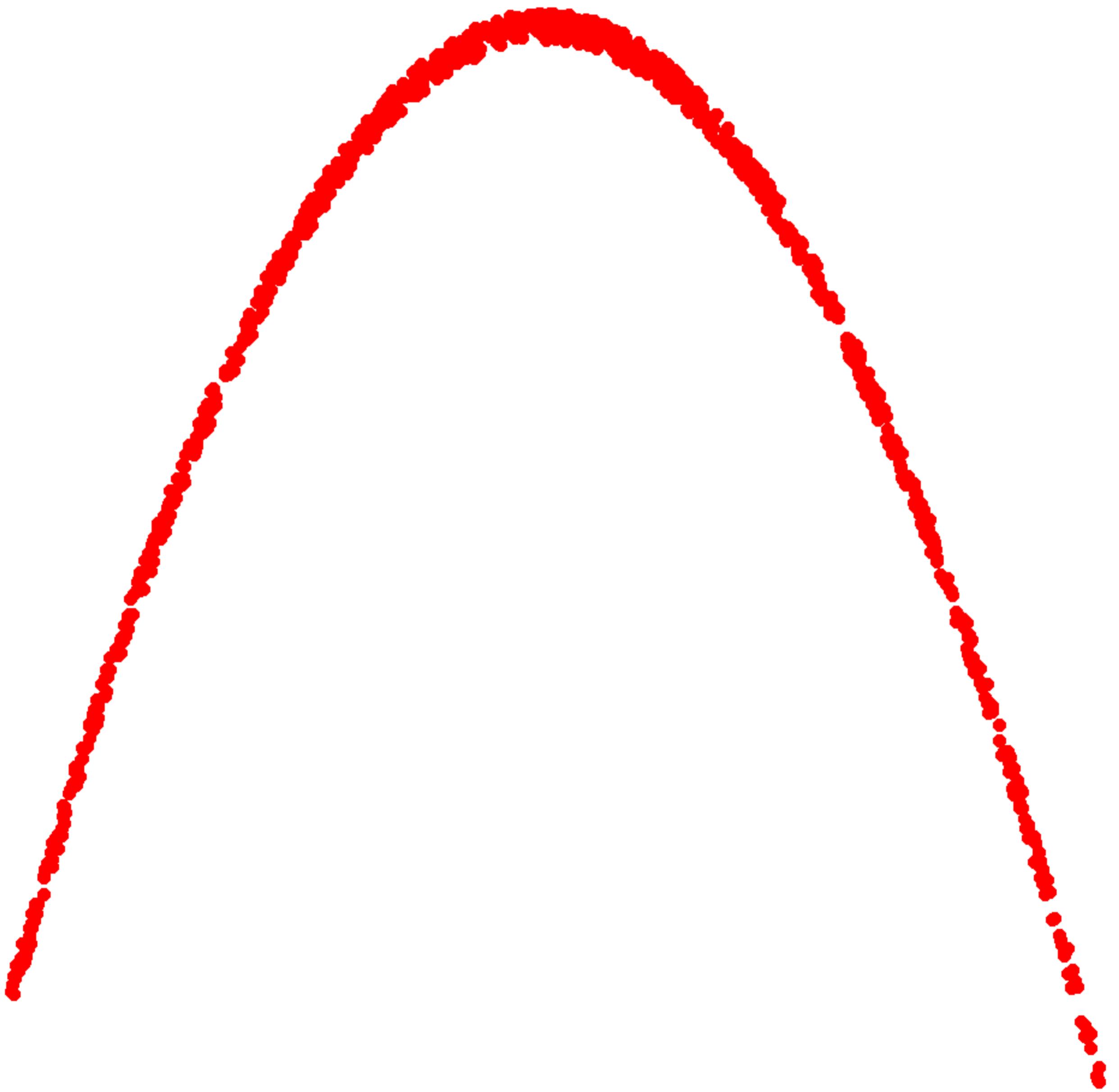
1



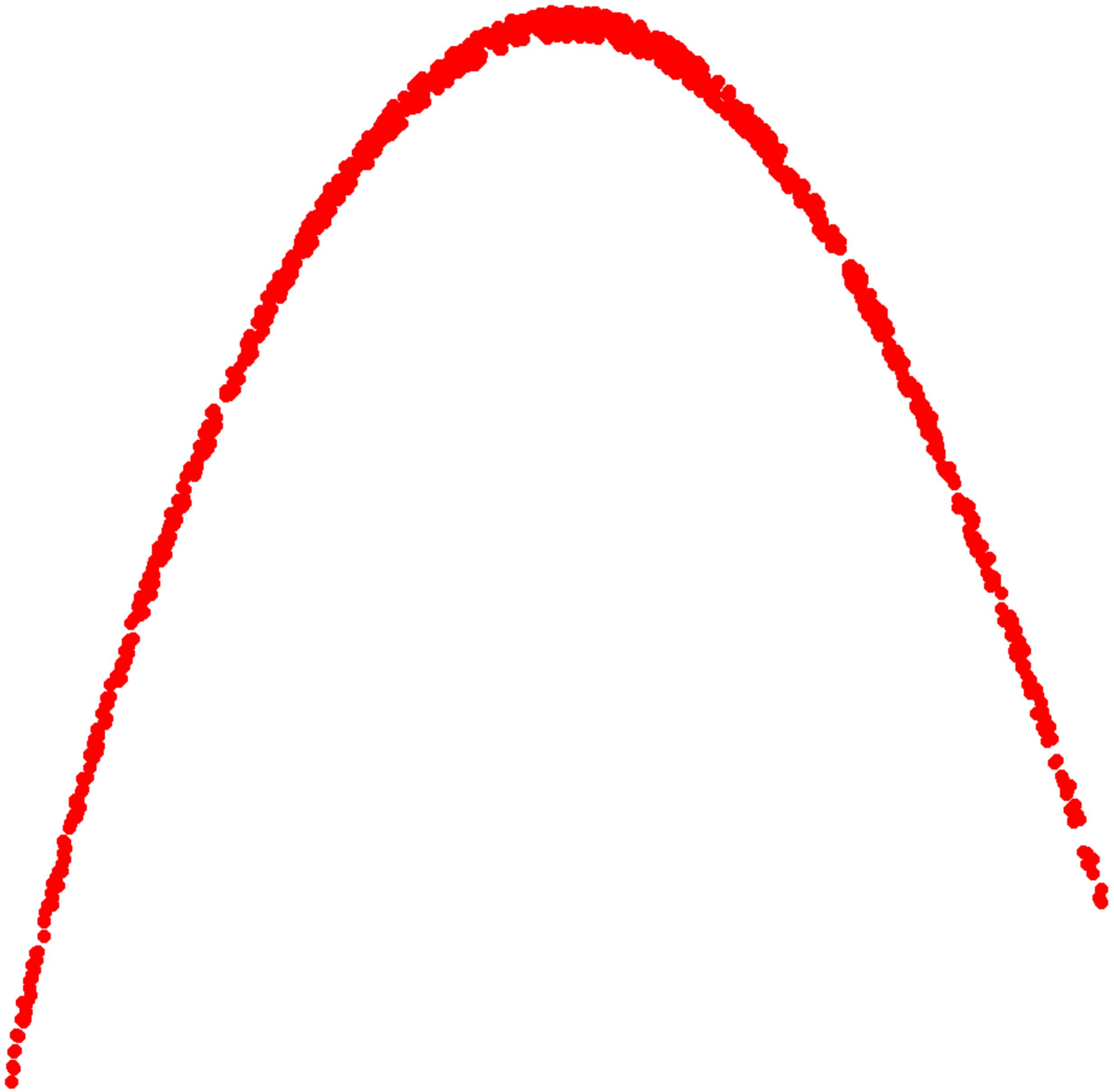
0.00166



0.00166



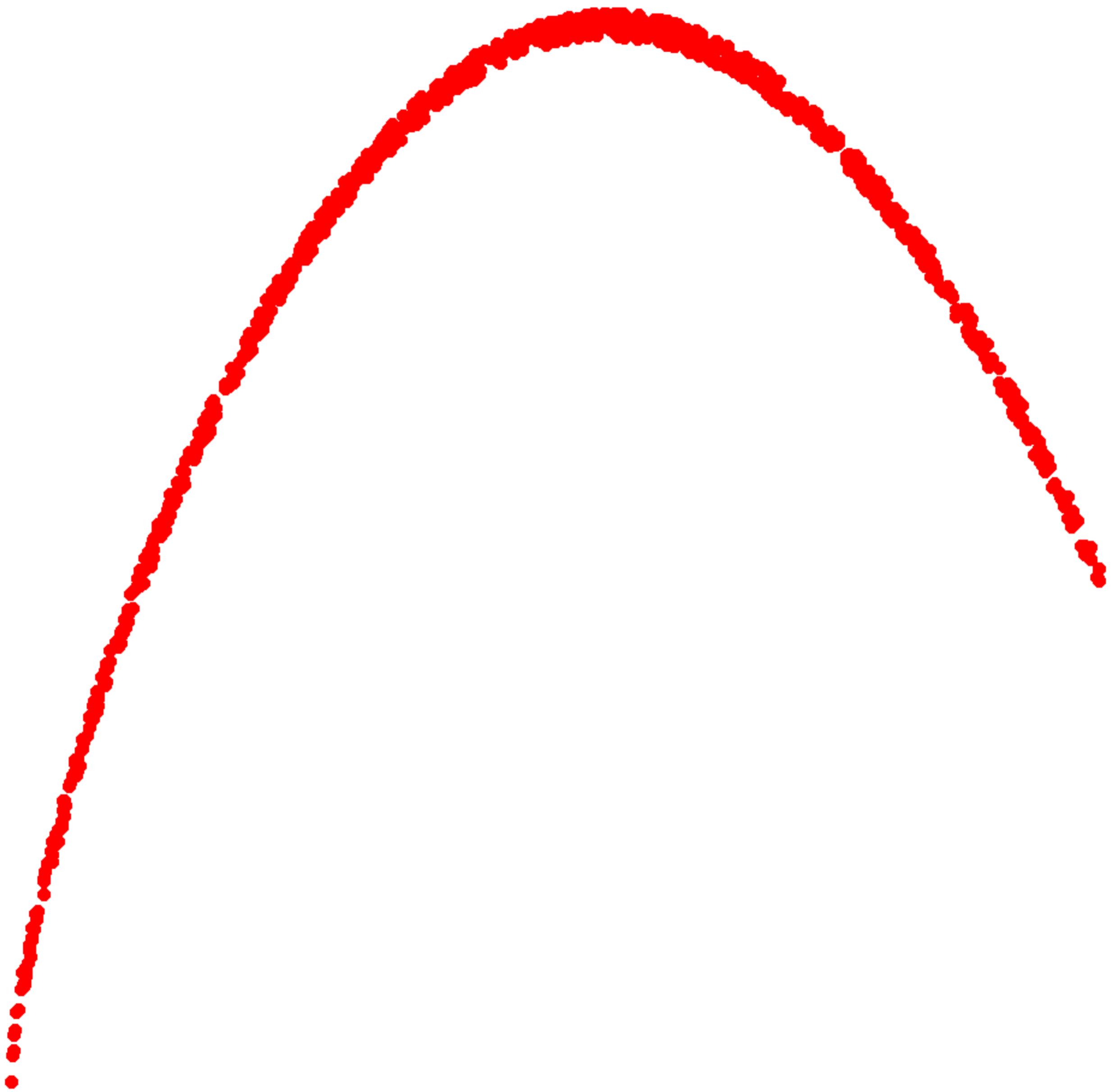
0.13287



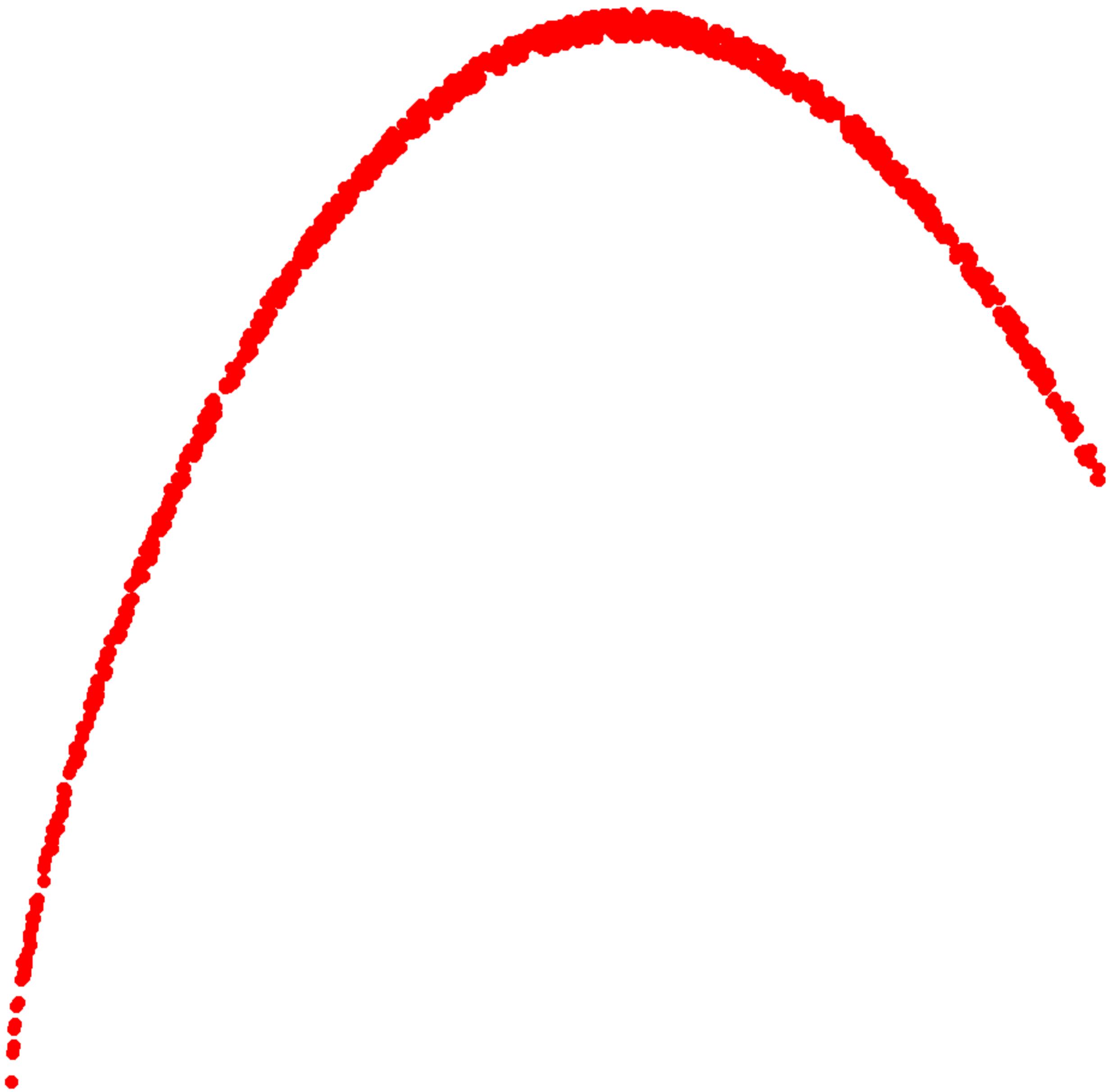
0.2412



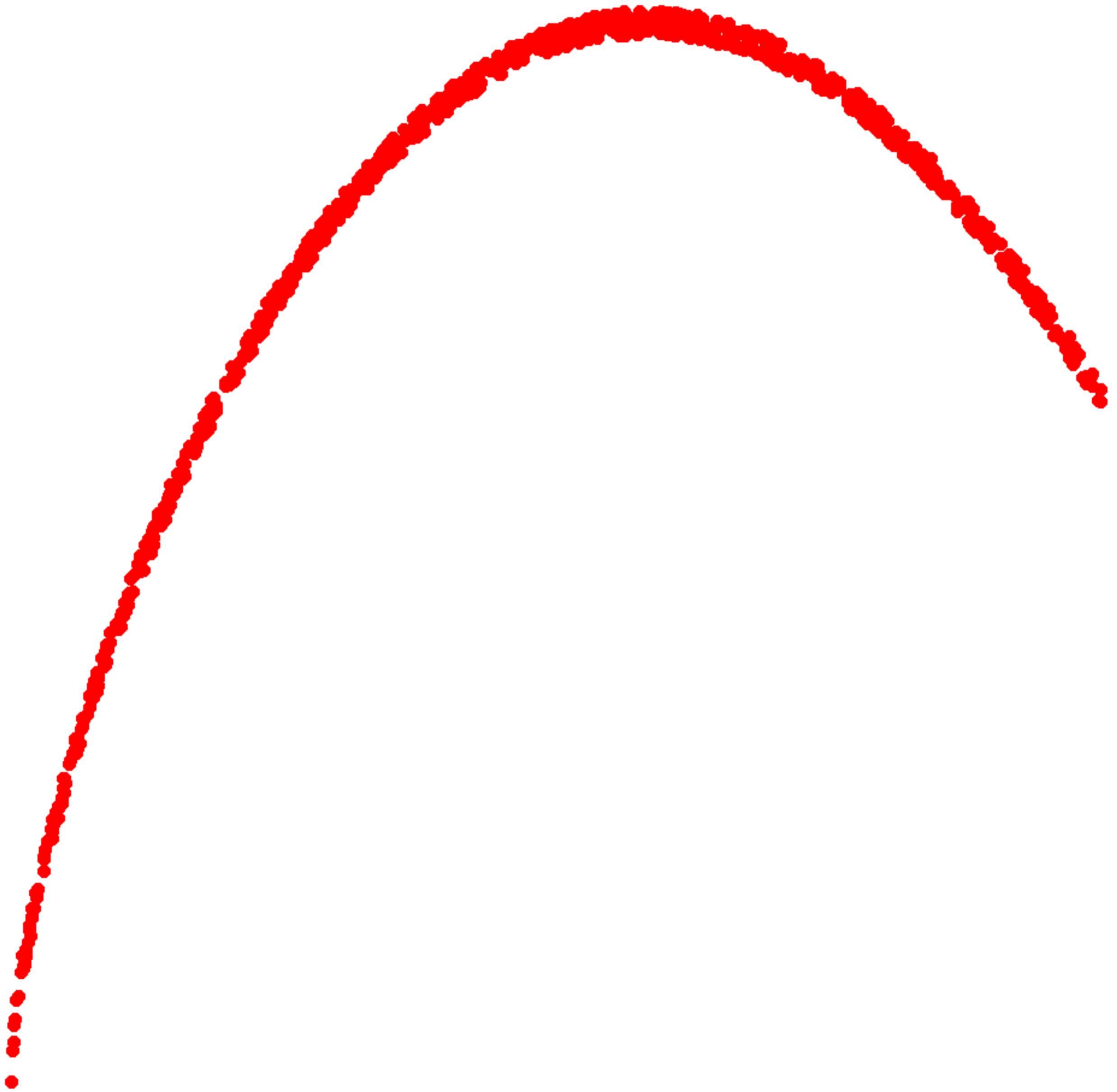
0.33339



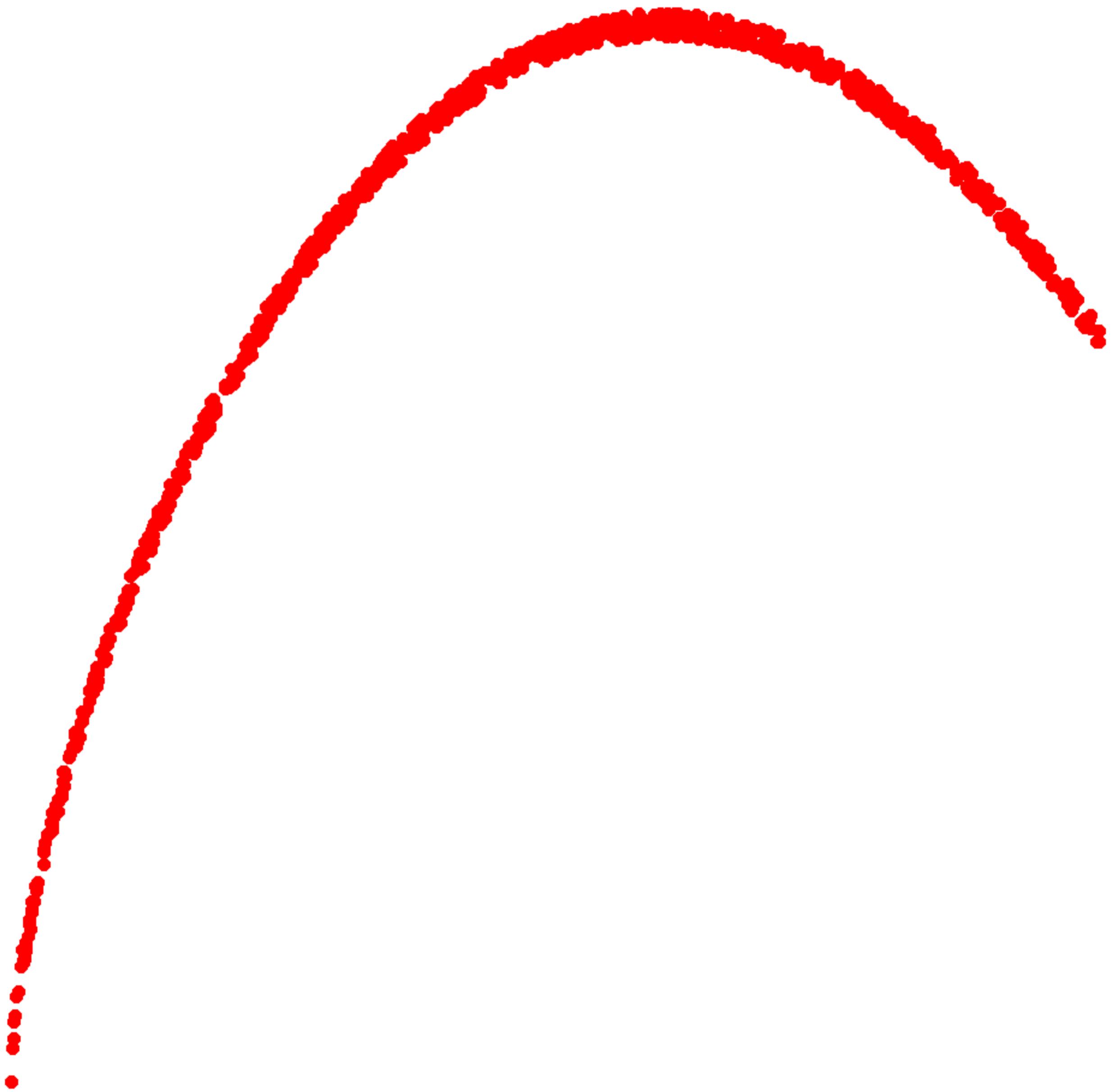
0.41284



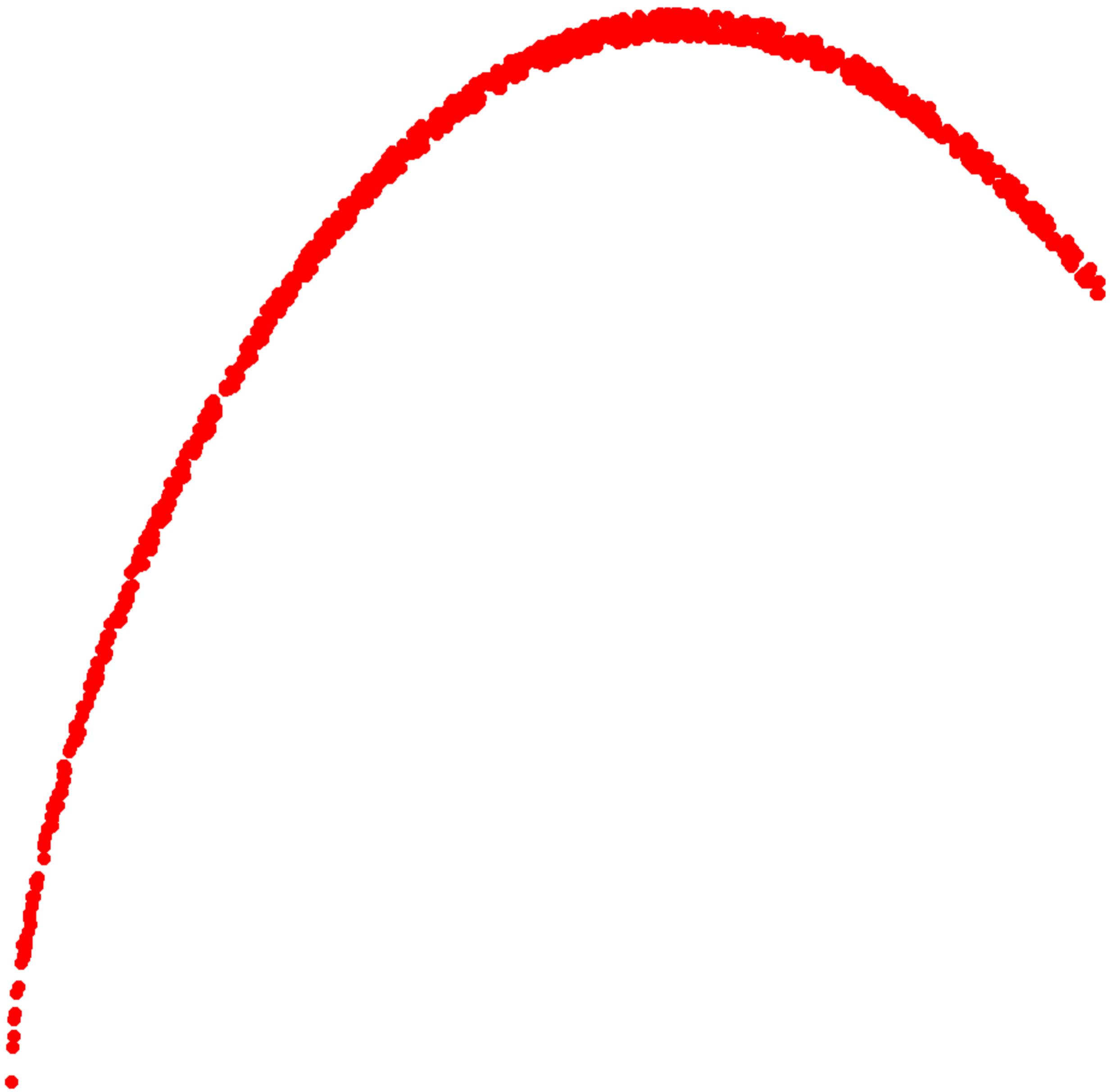
0.48154



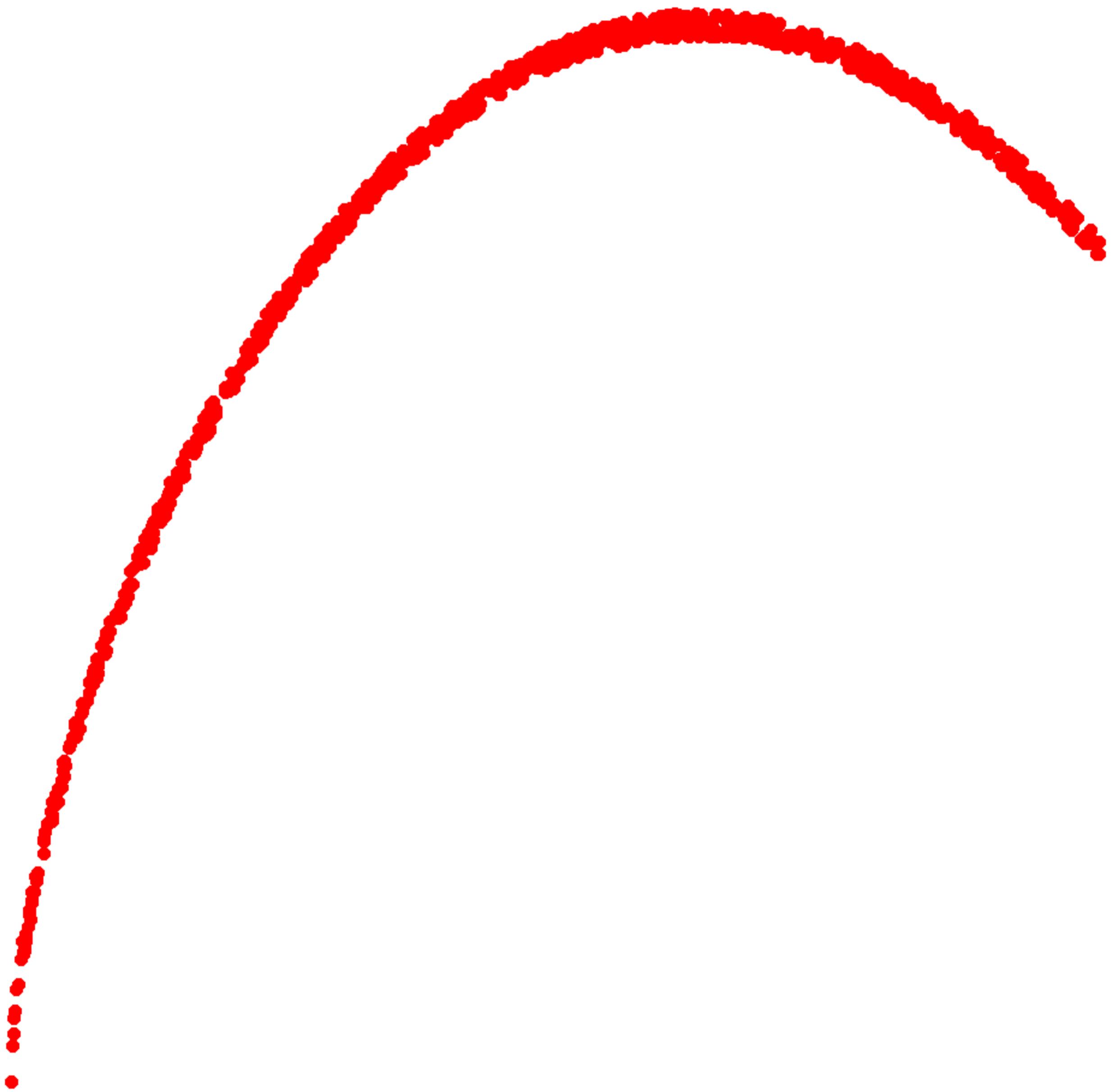
0.54395



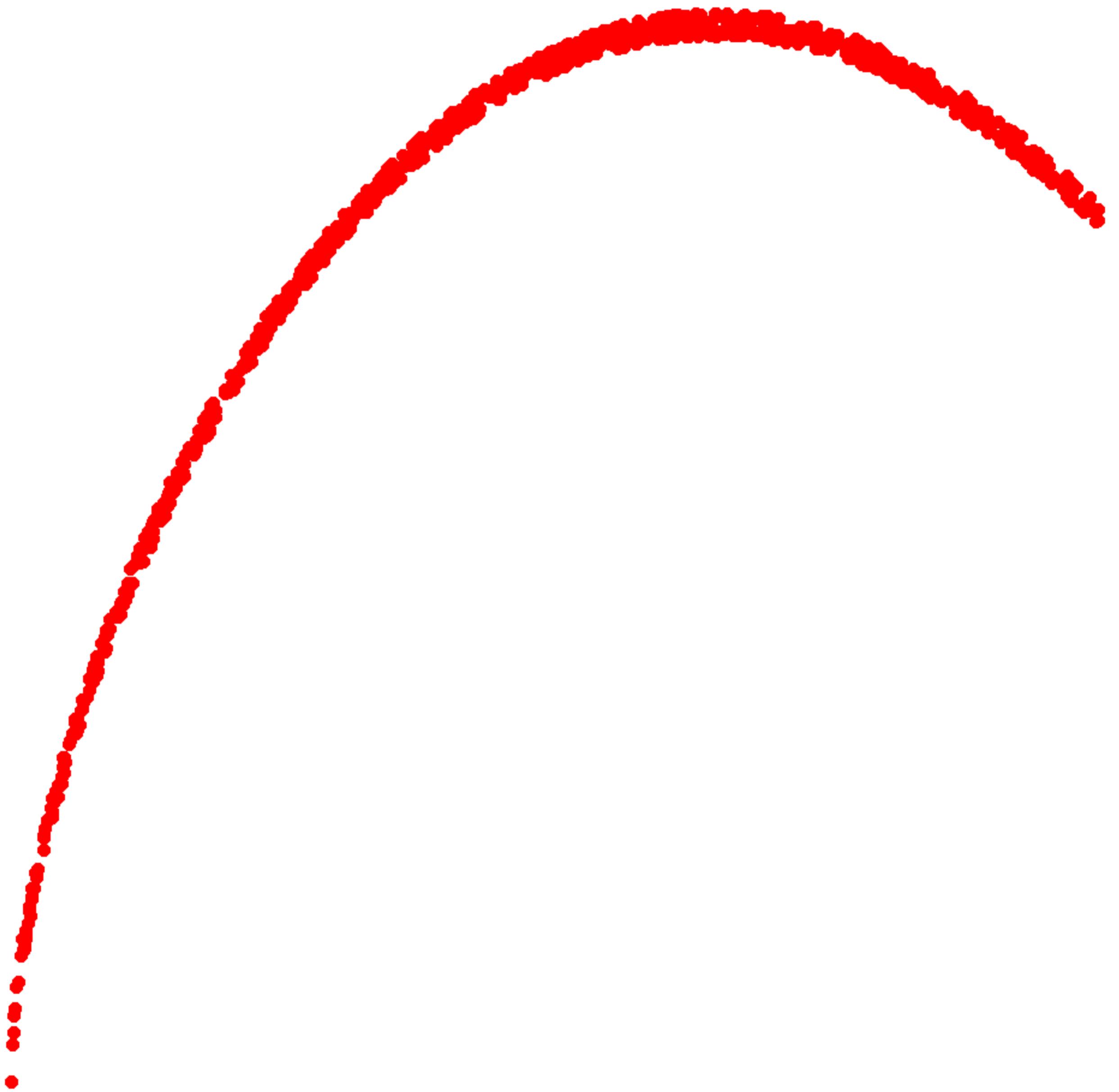
0.59818



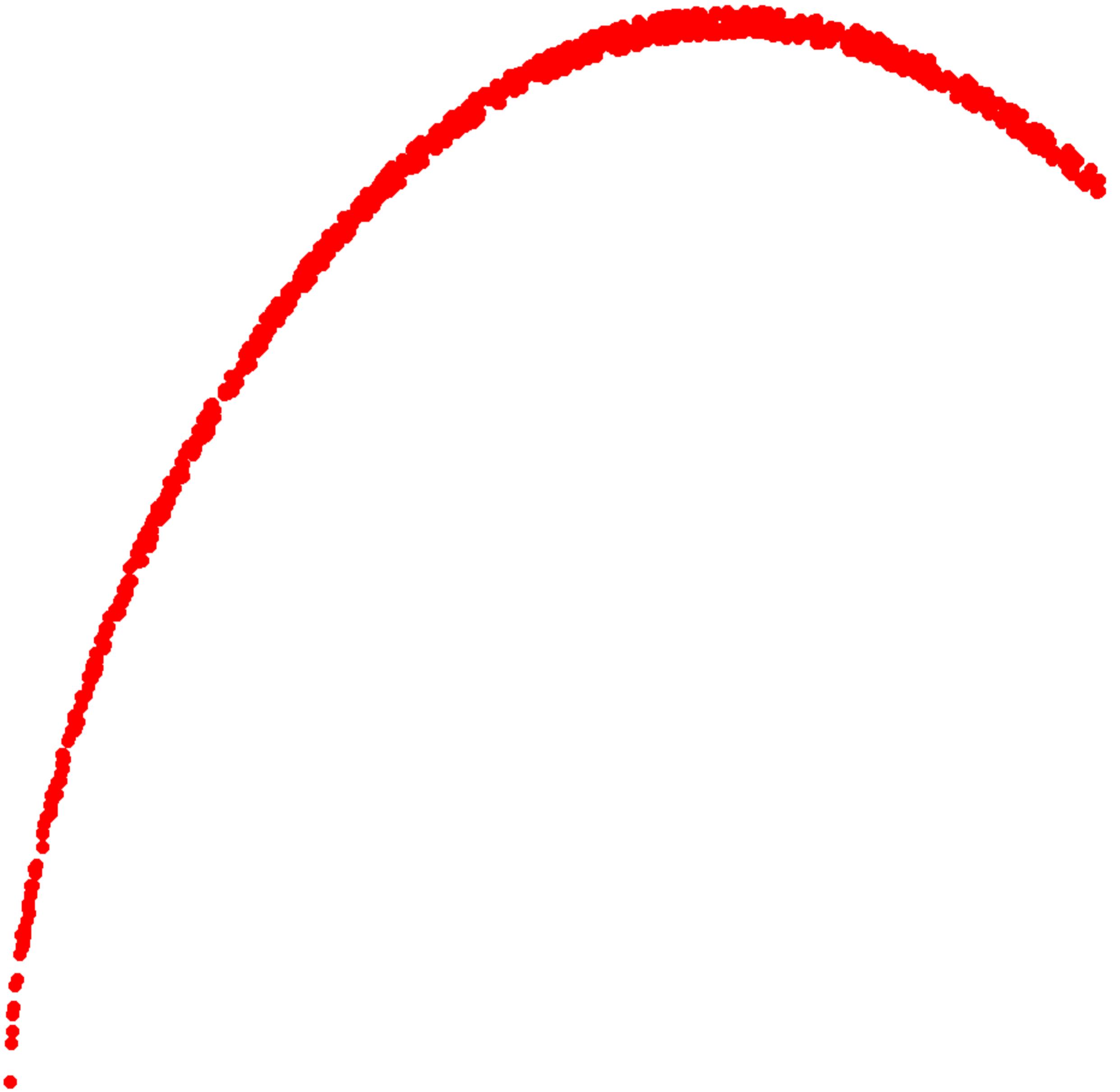
0.64592



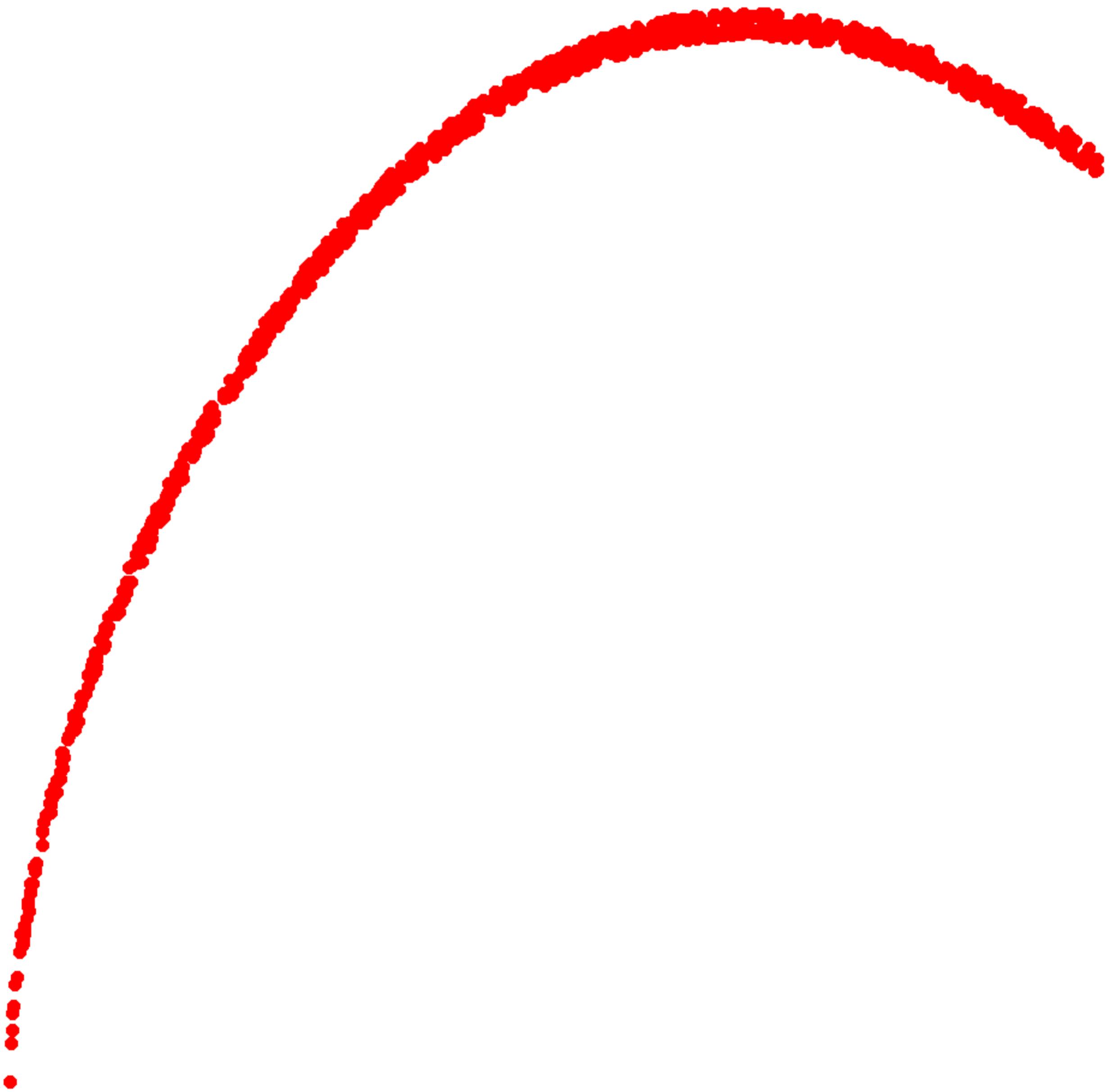
0.68802



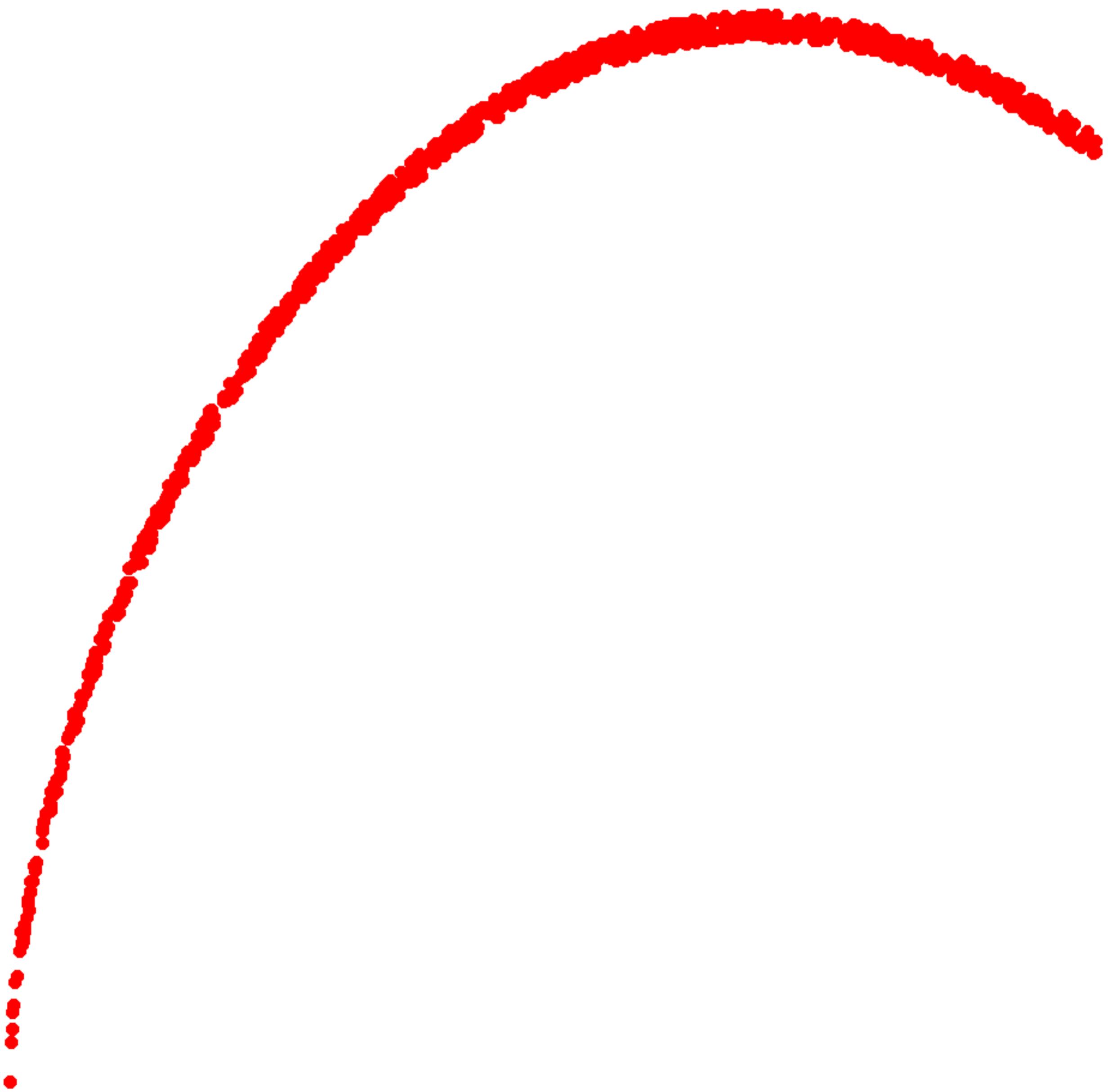
0.72696



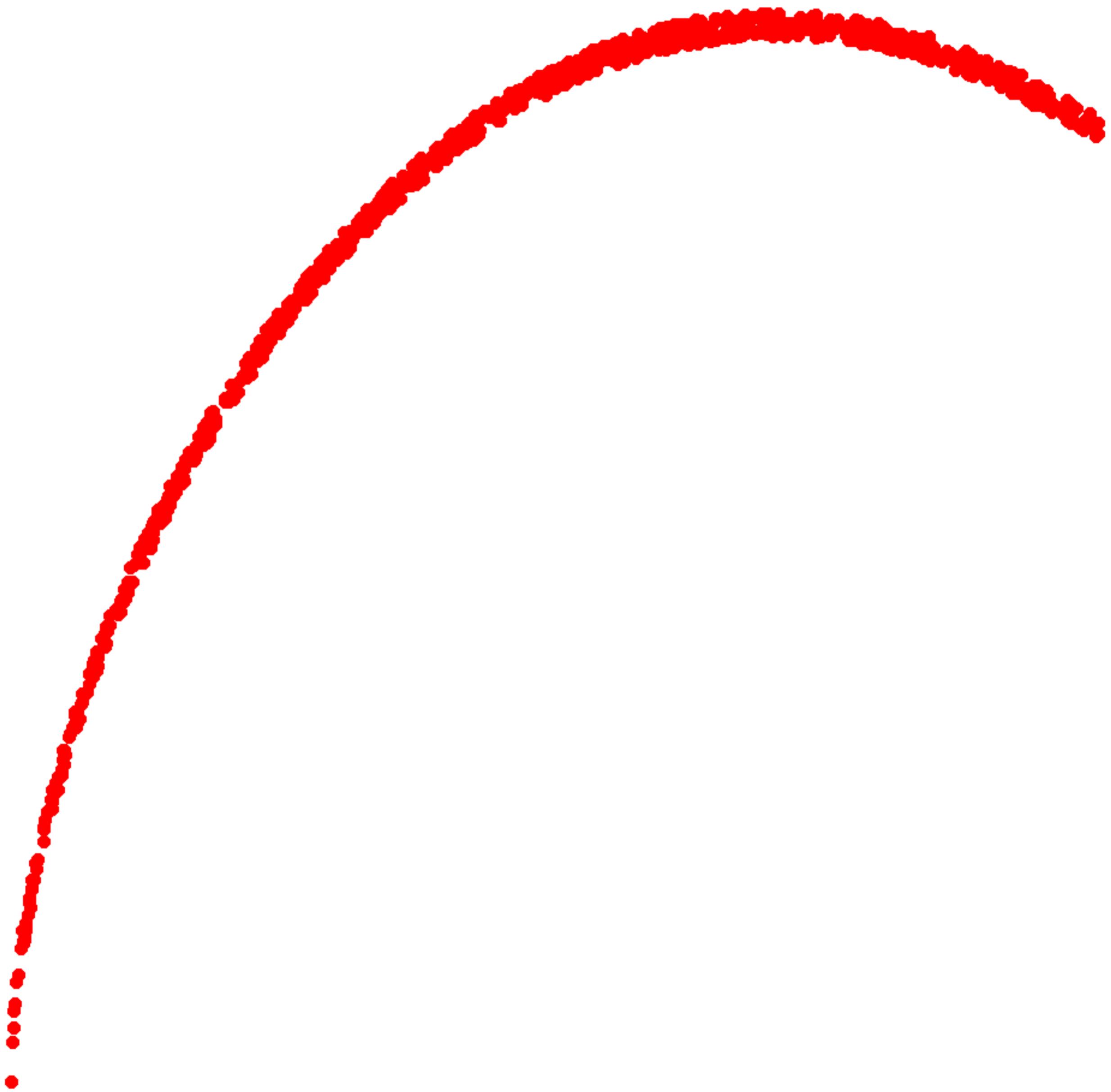
0.76222



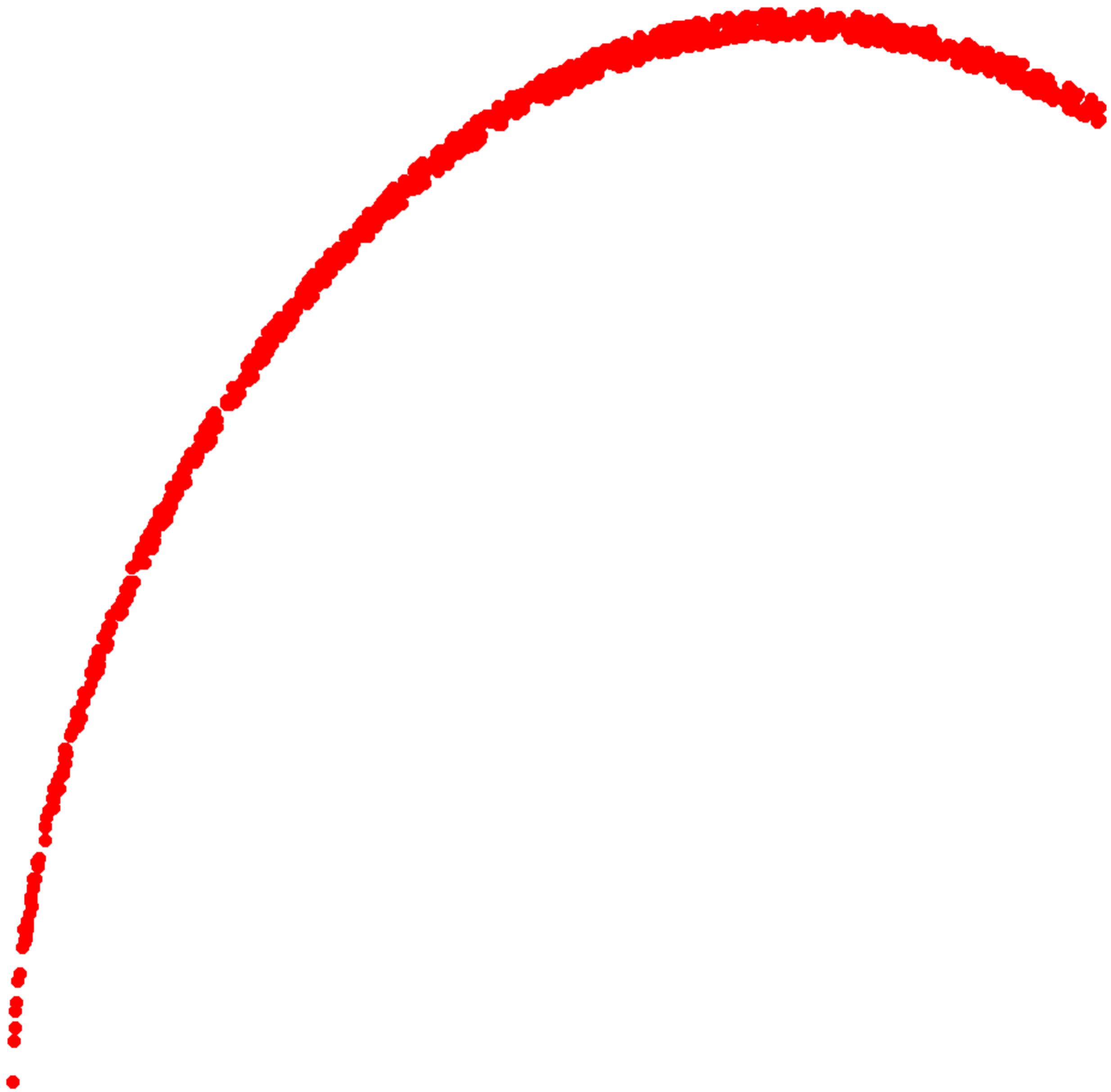
0.79396



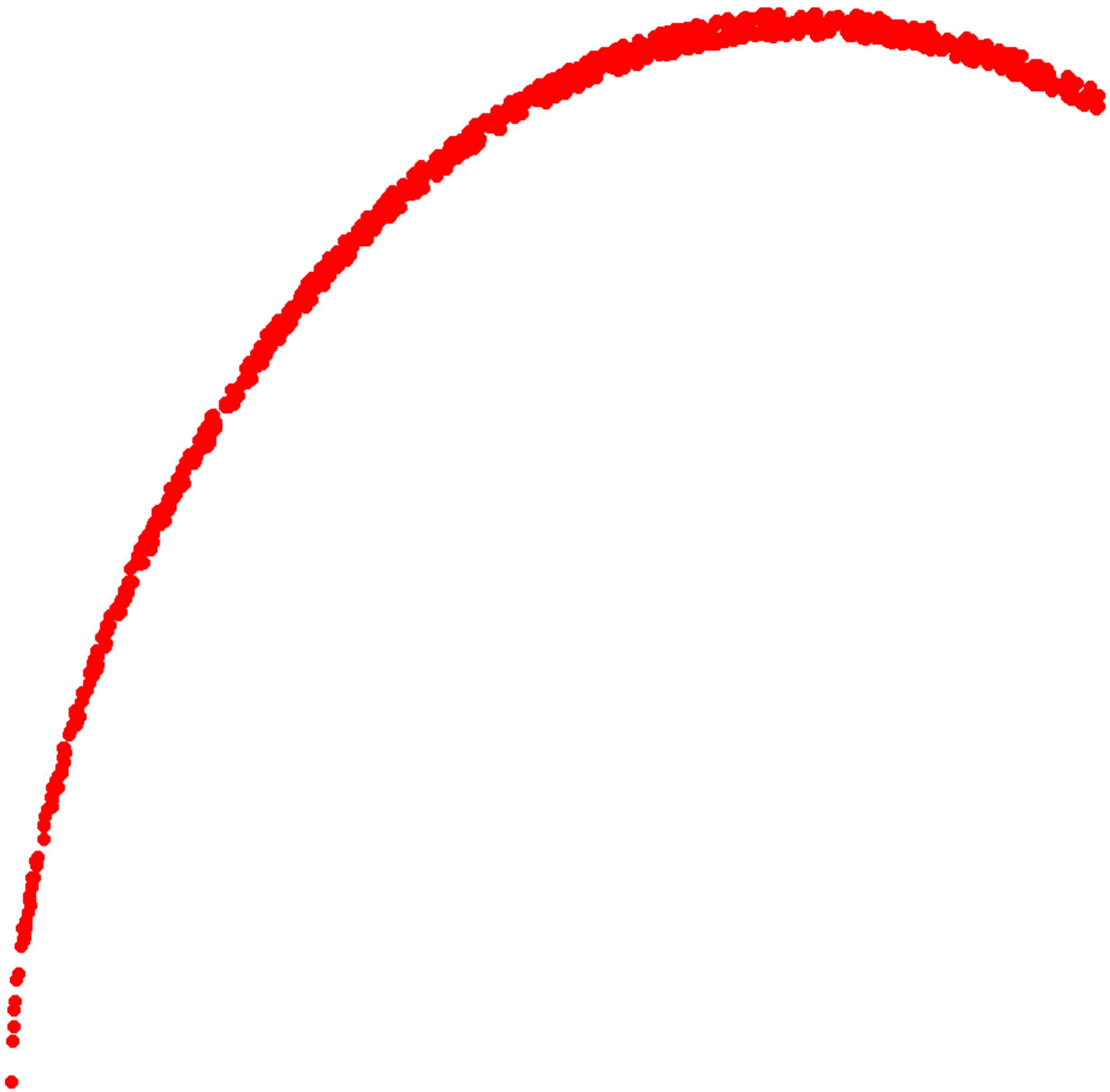
0.82219



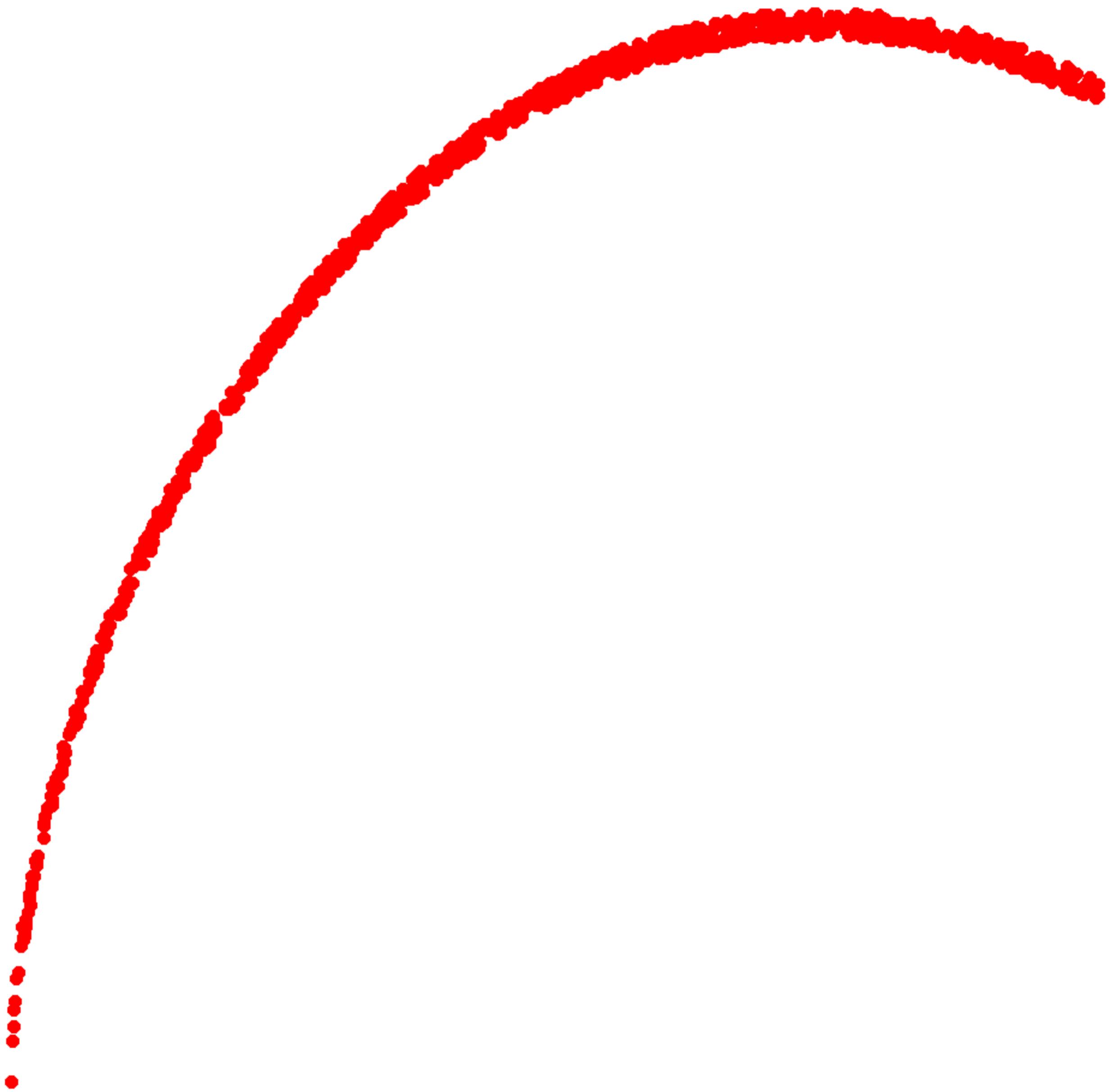
0.8472



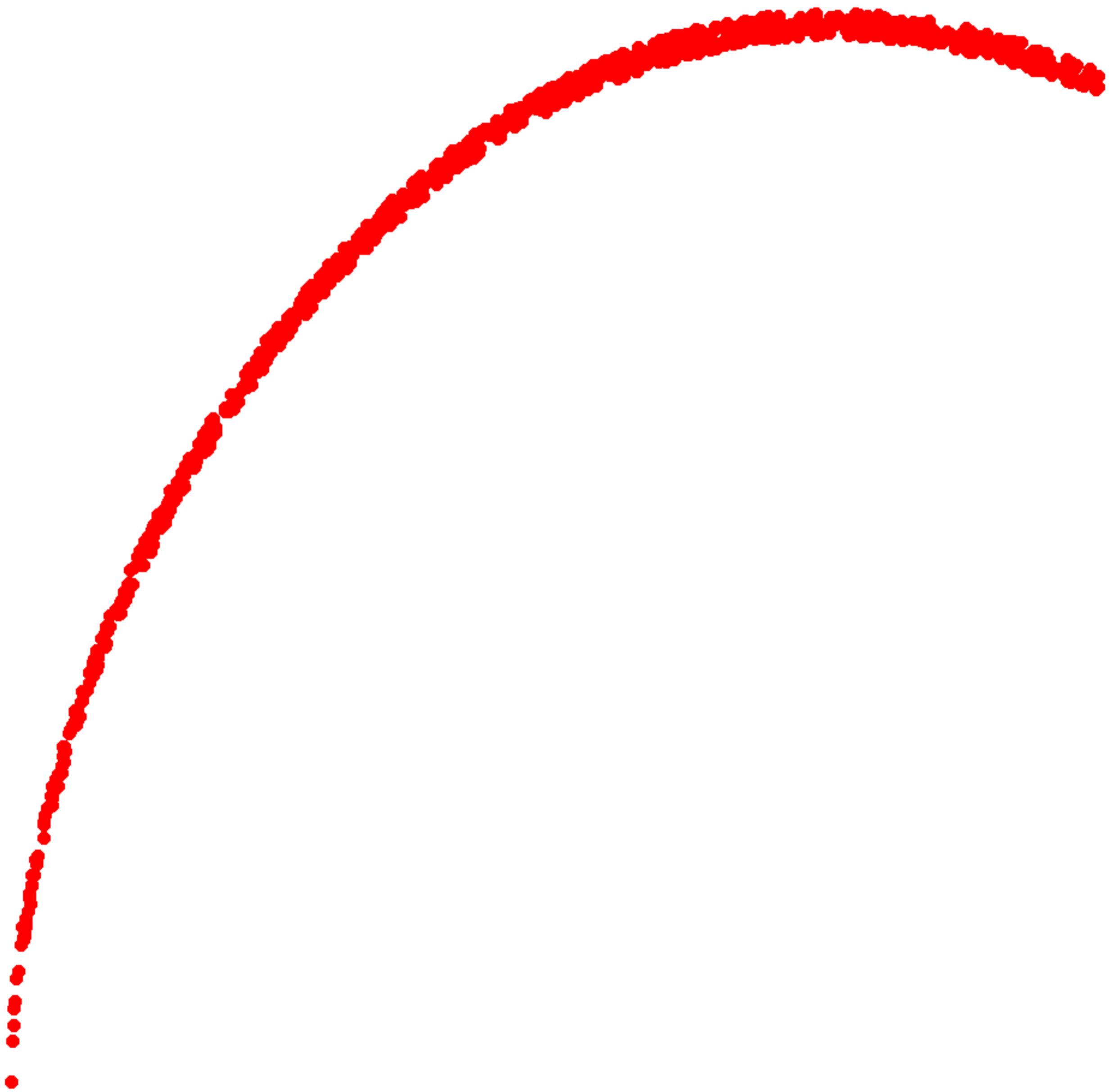
0.86999



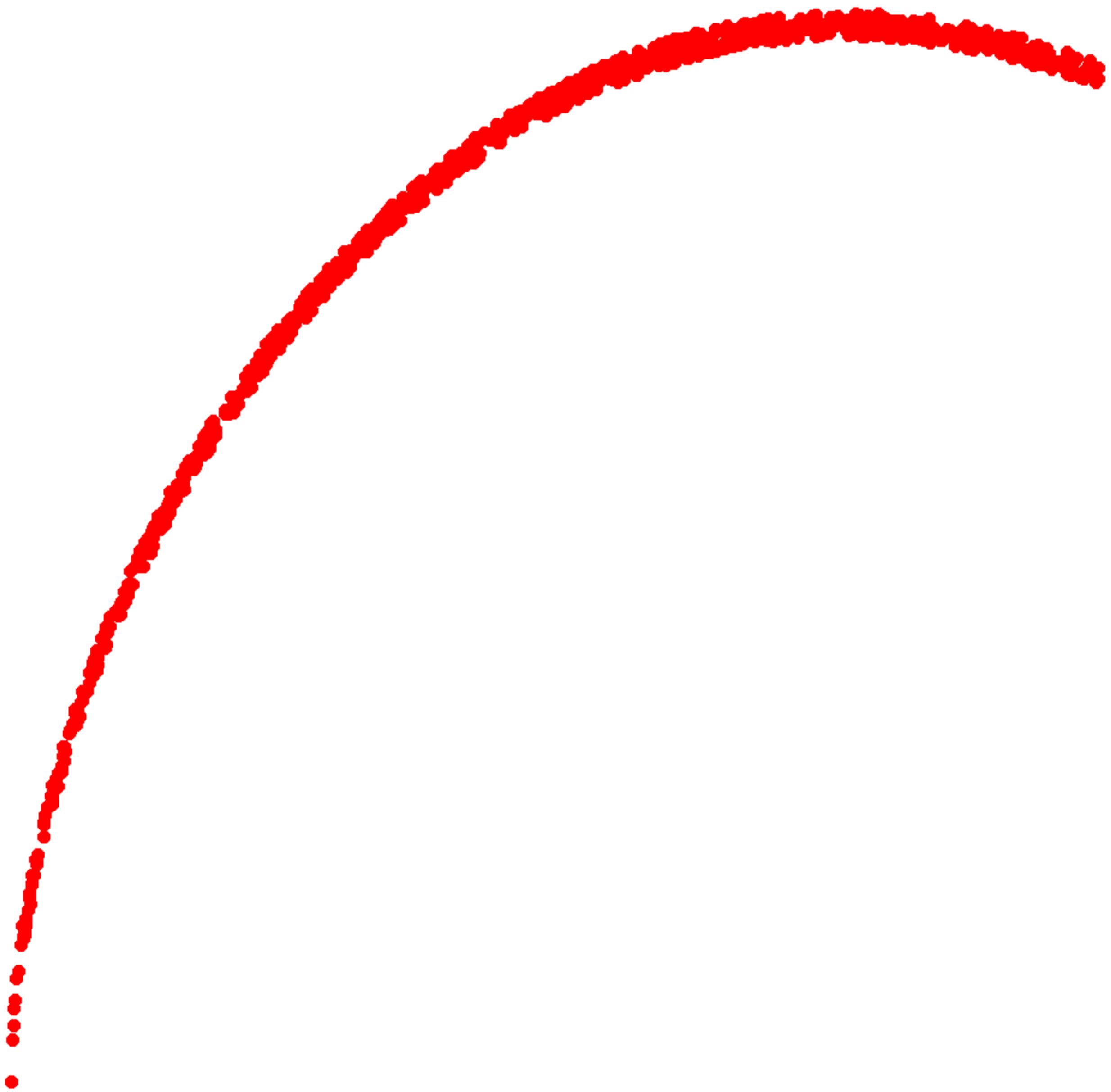
0.89045



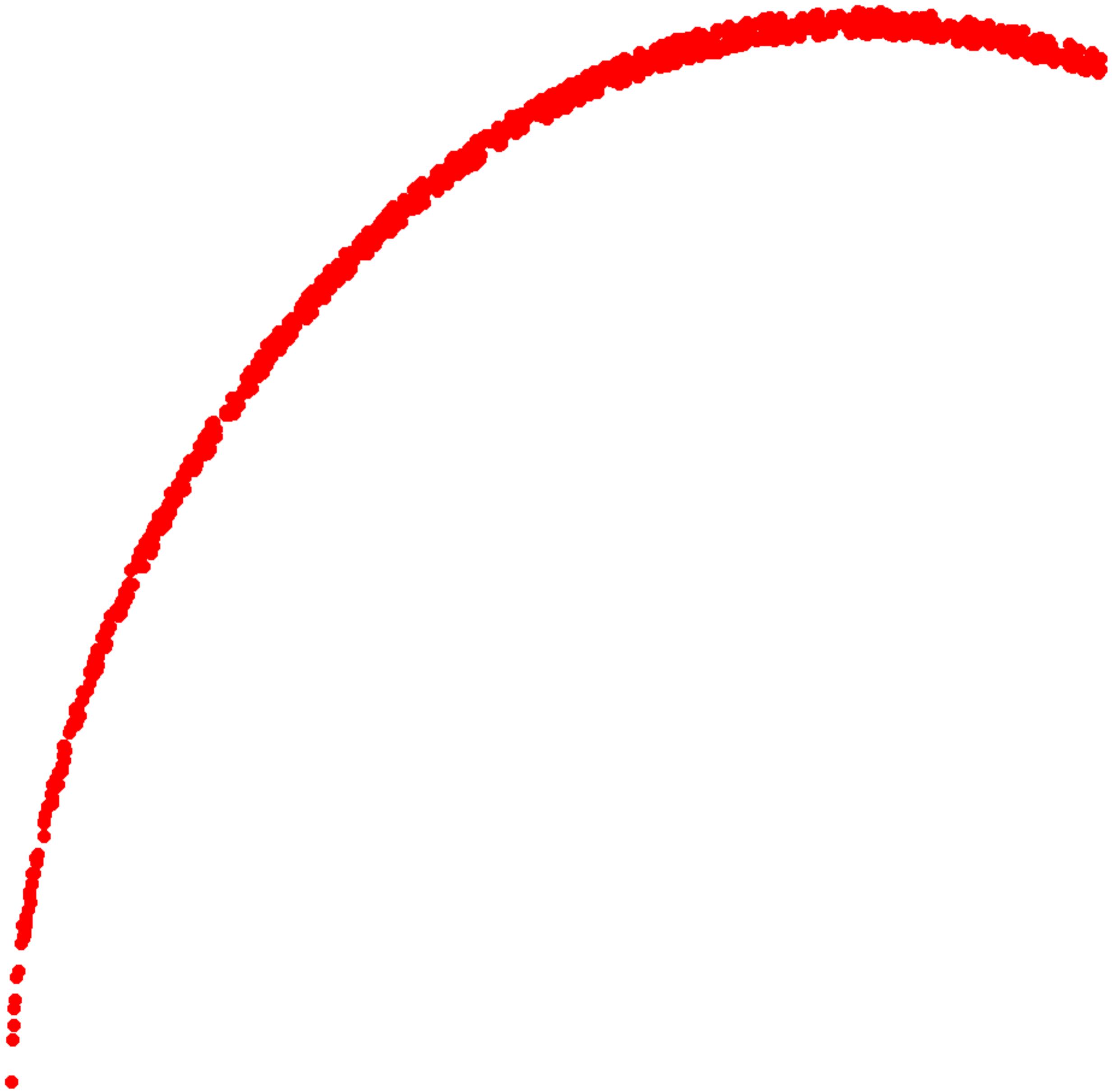
0.90848



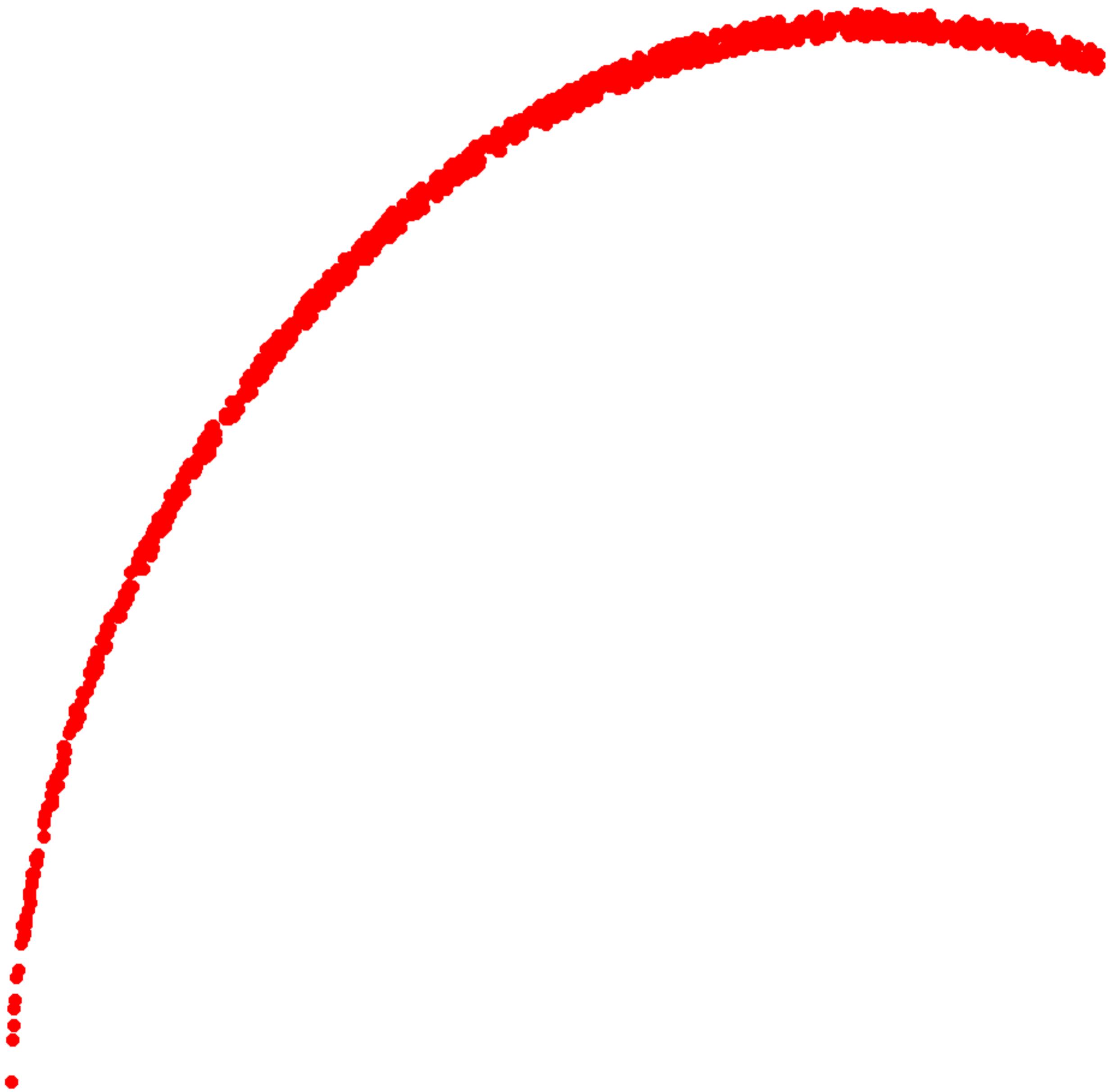
0.92366



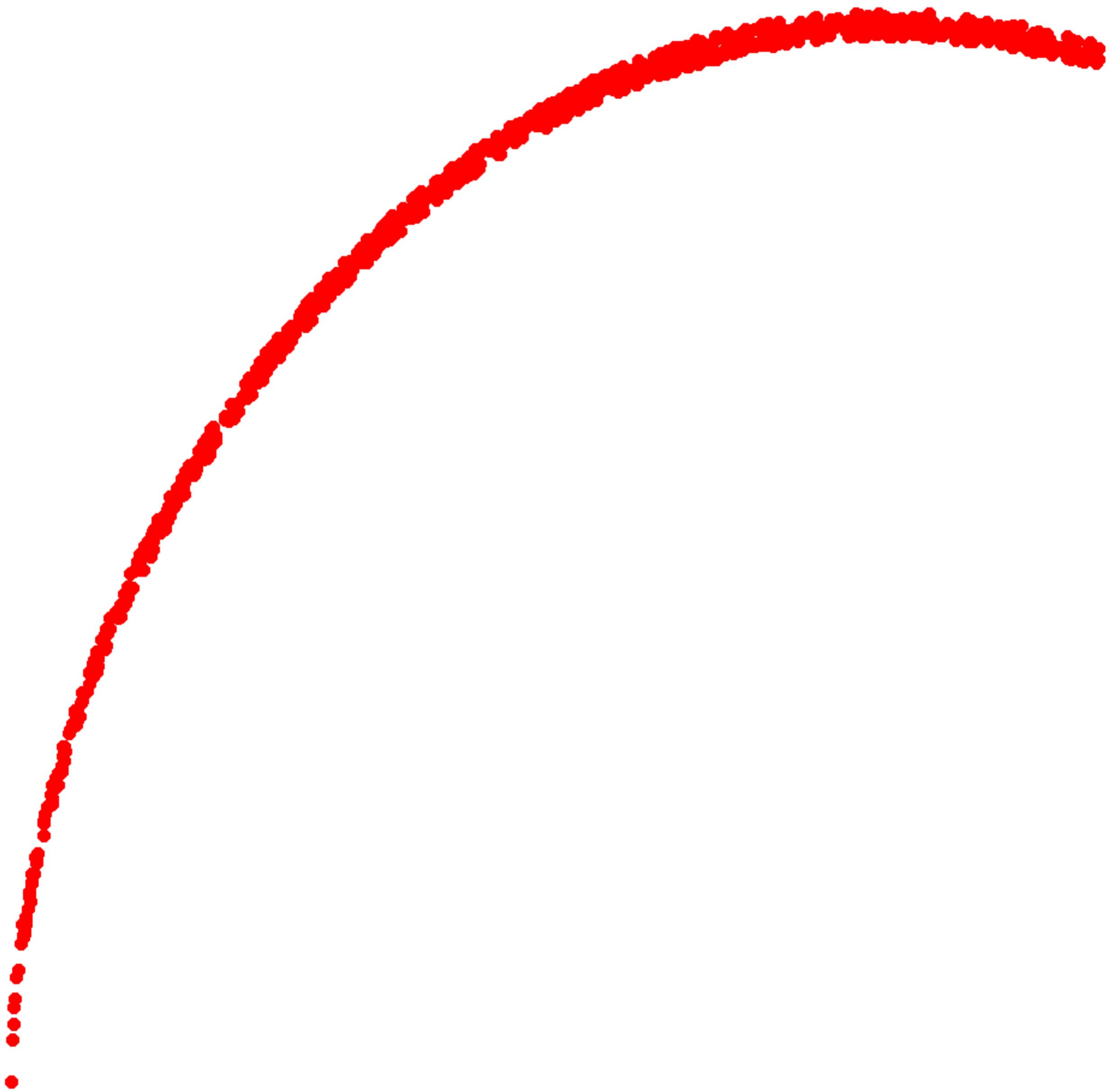
0.93611



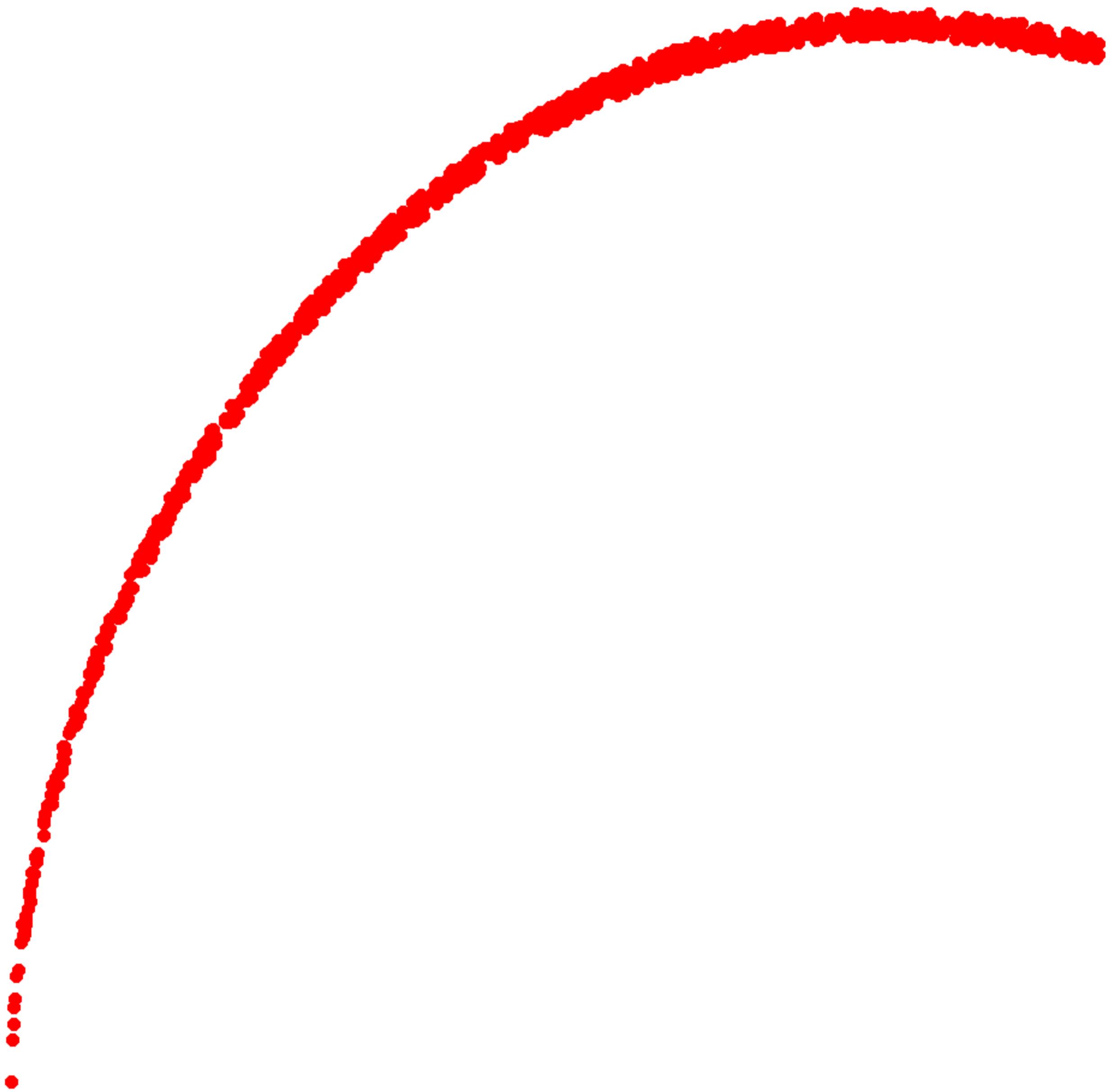
0.94696



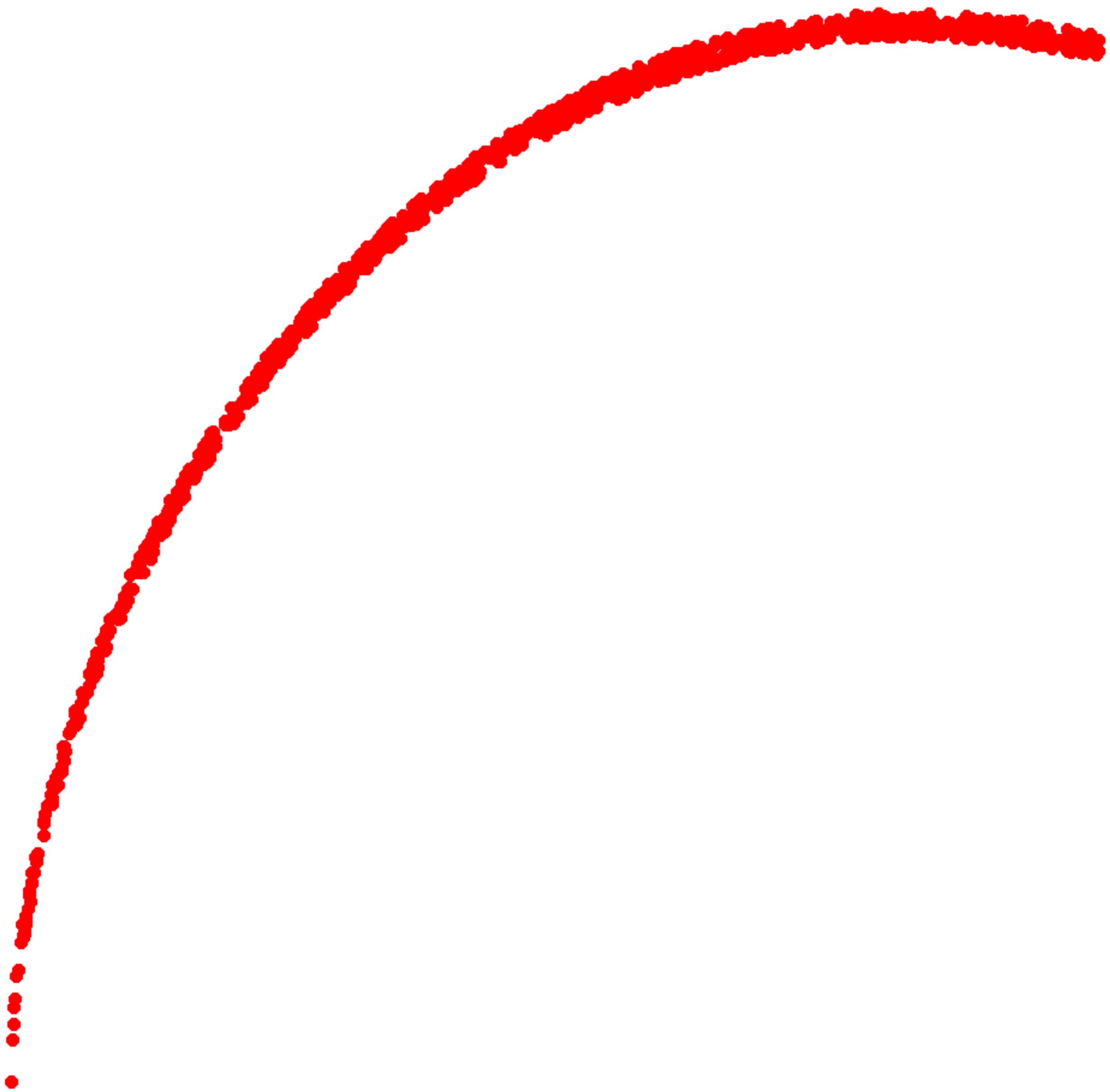
0.95587



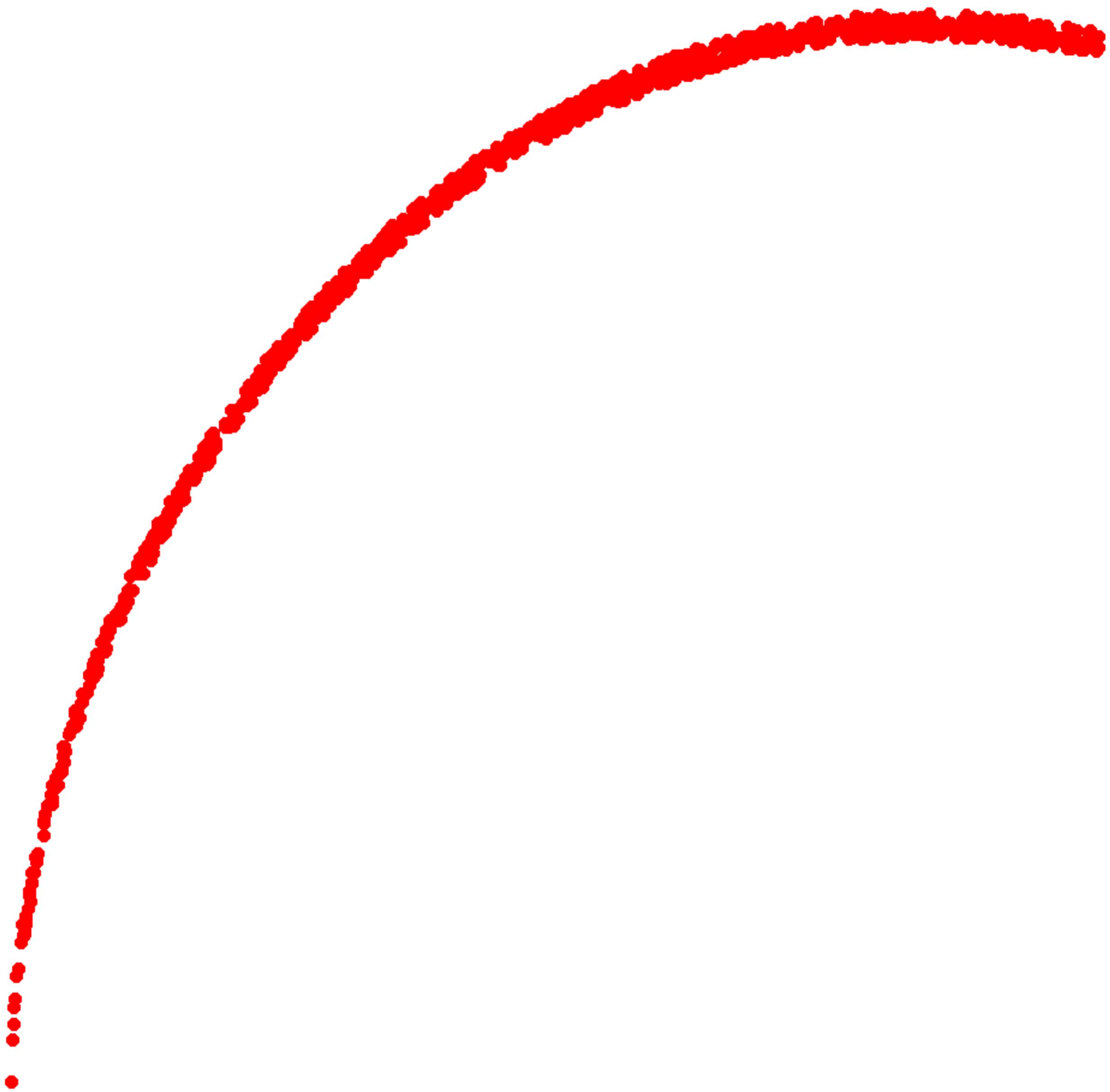
0.96325



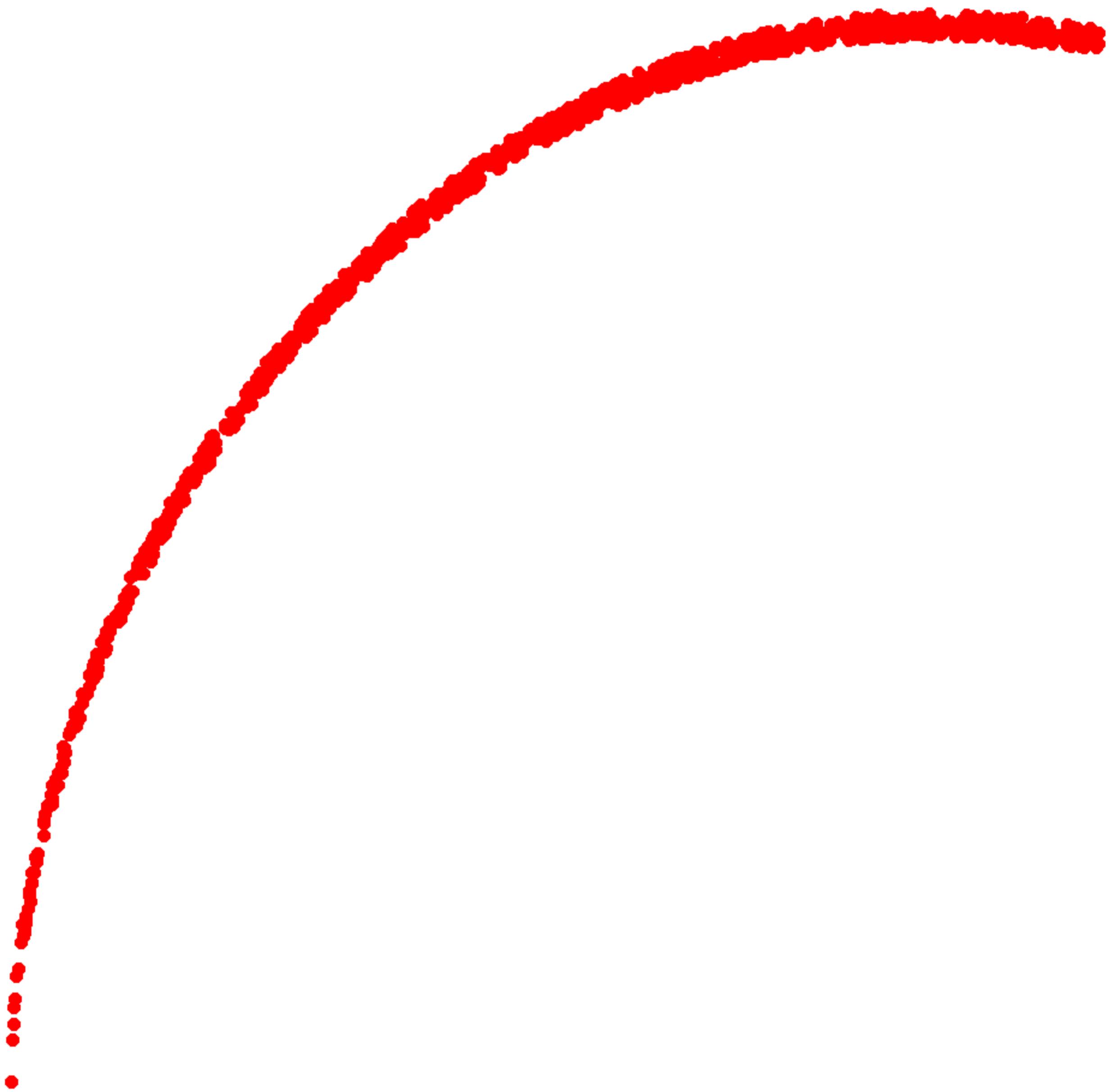
0.96957



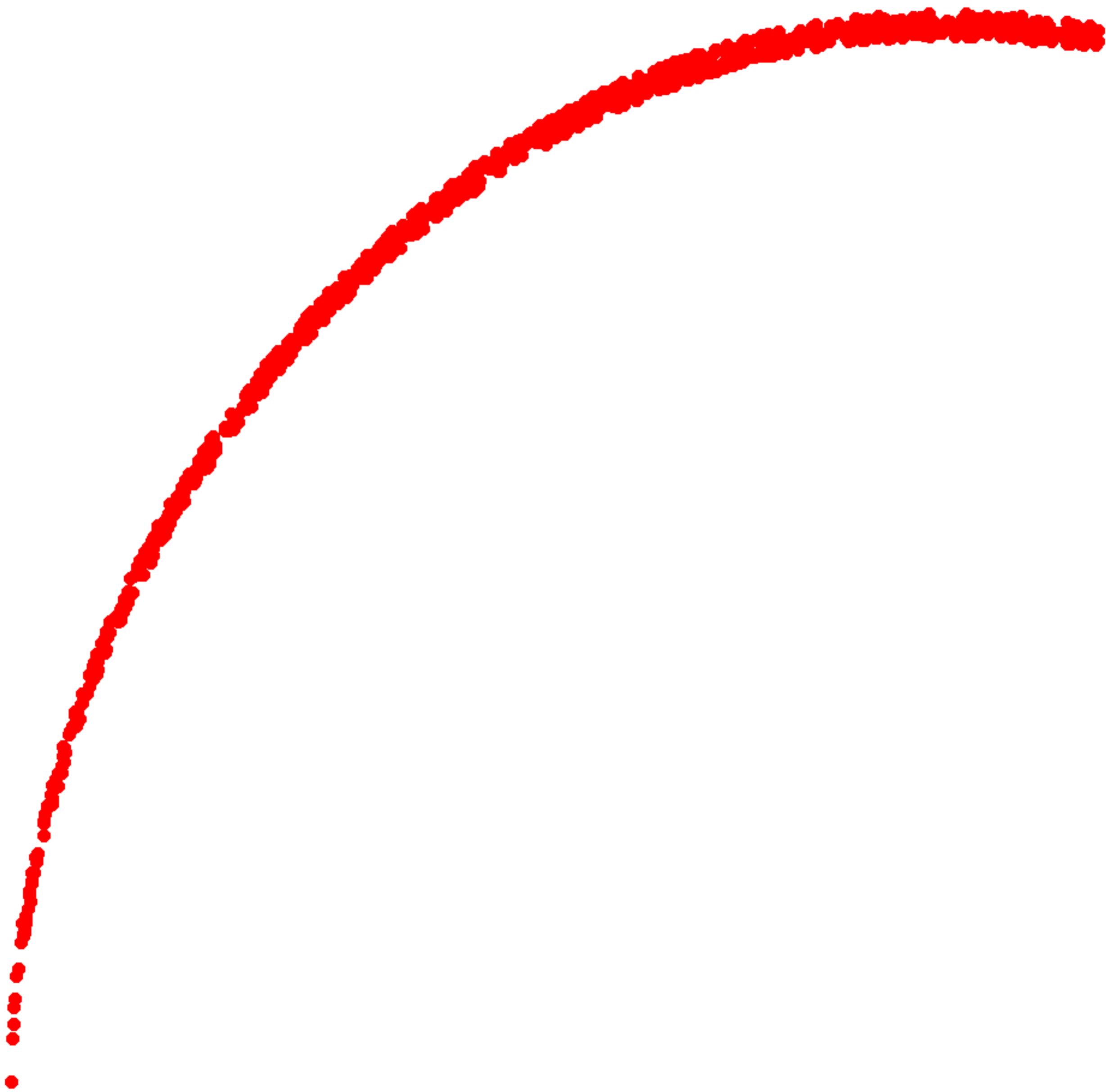
0.97484



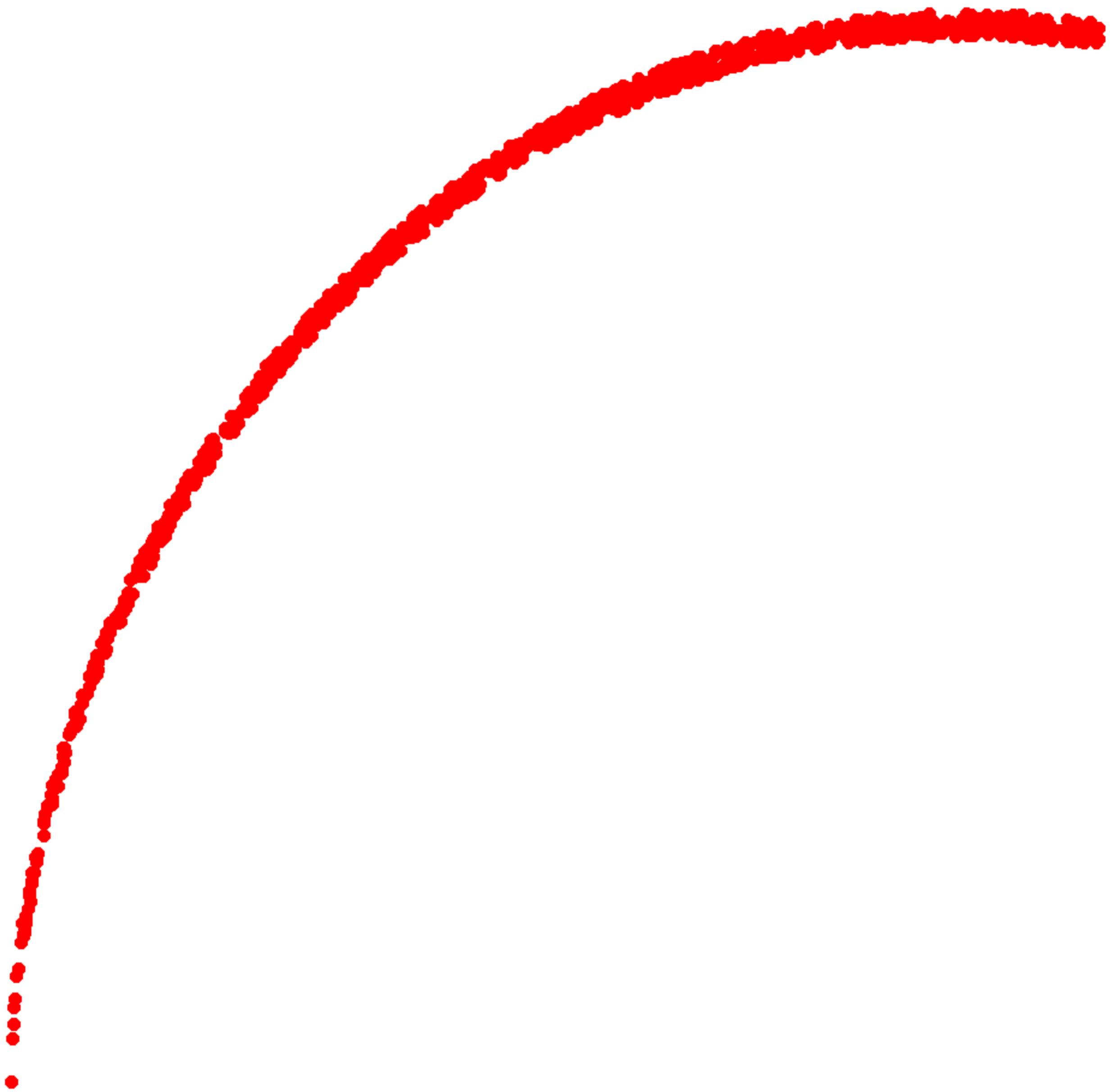
0.97918



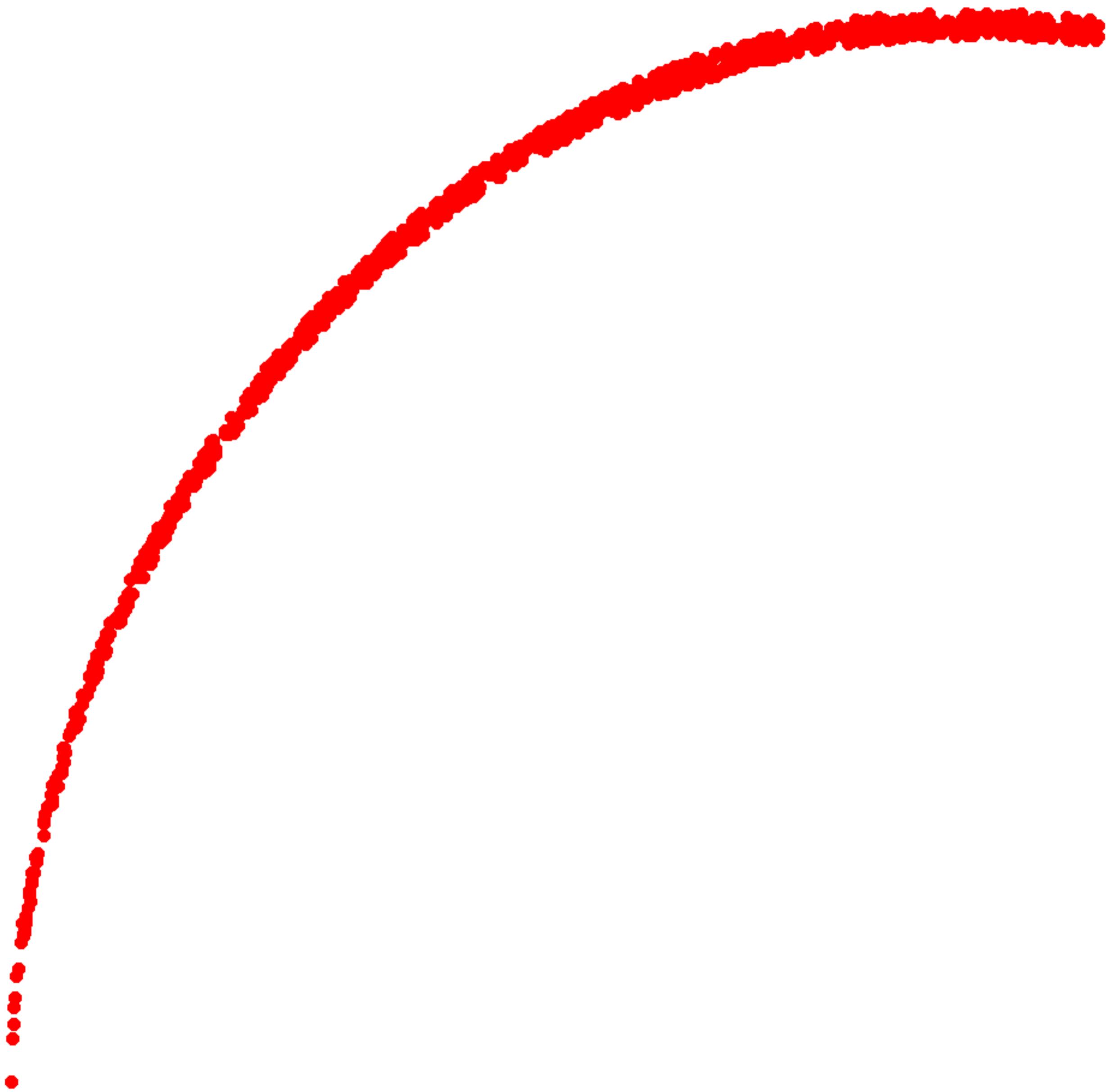
0.98285



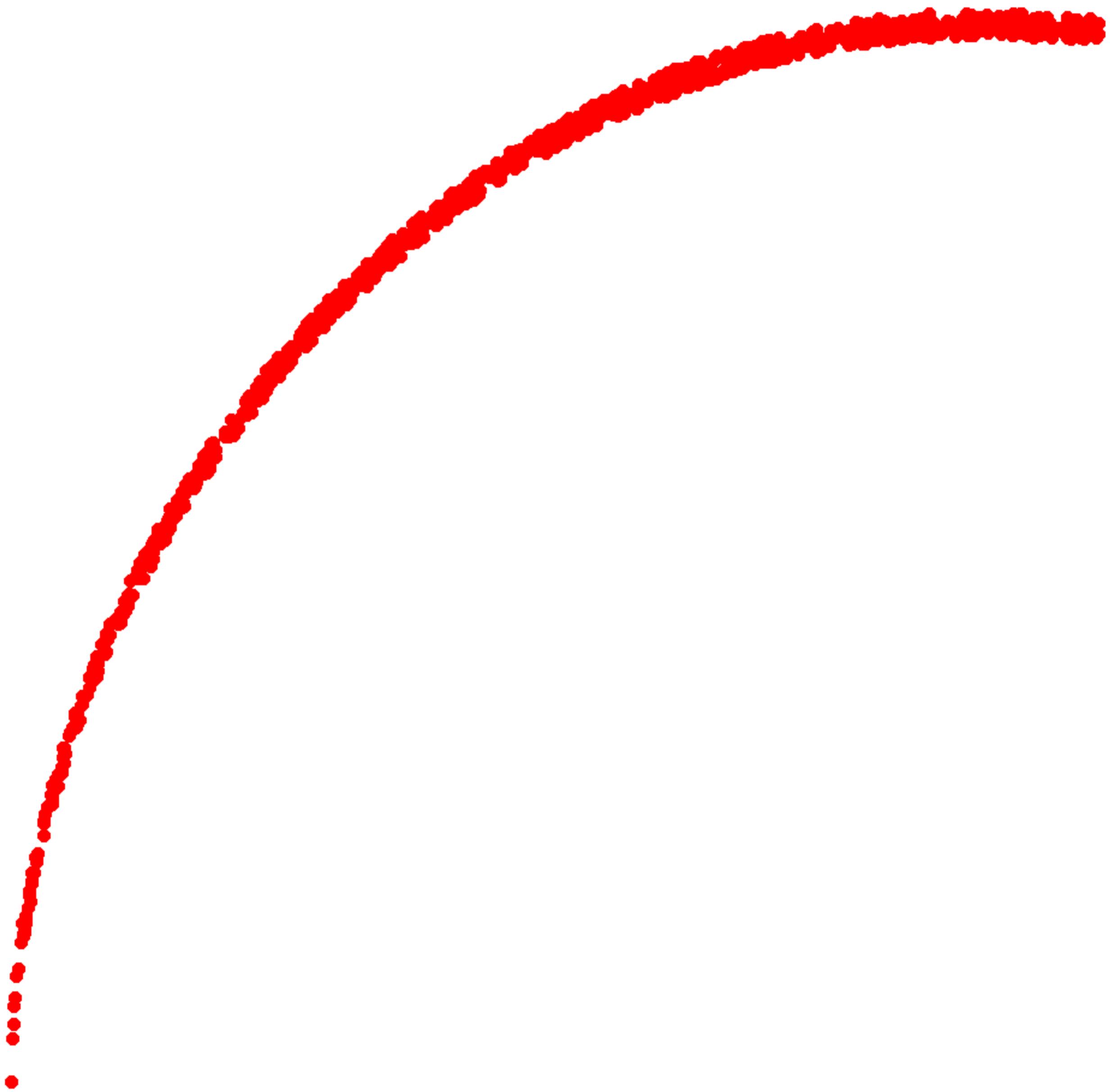
0.98581



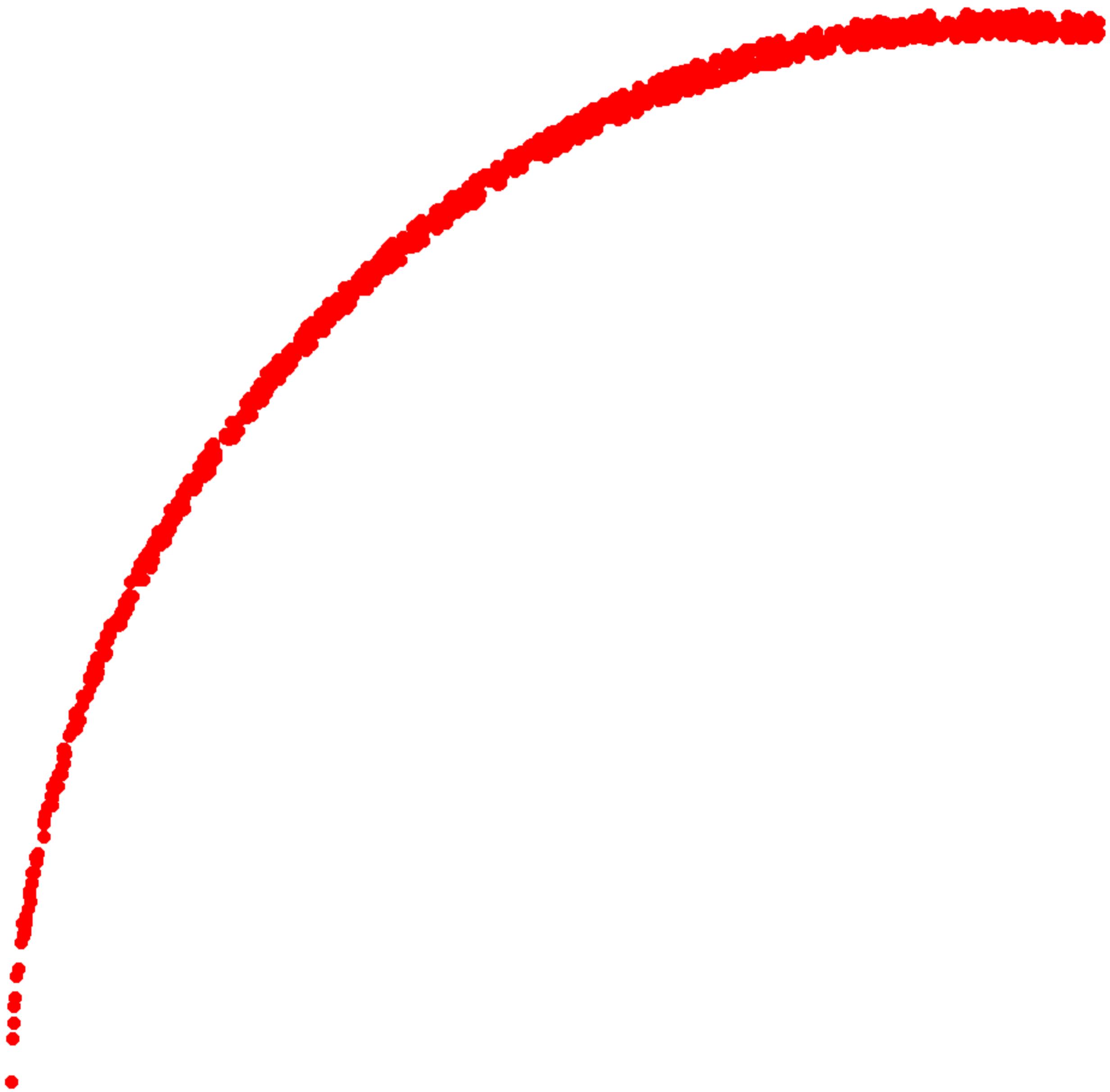
0.98824



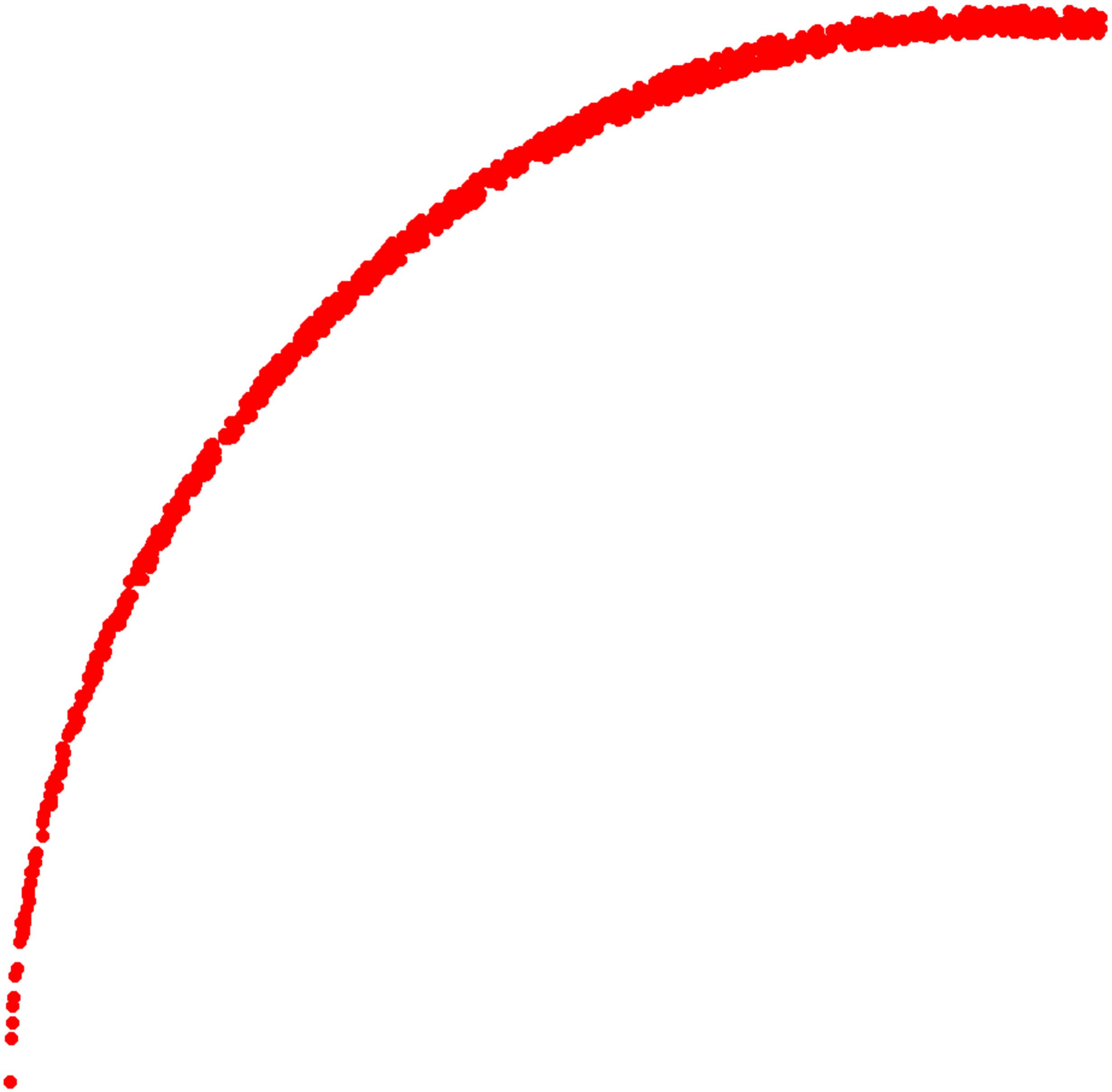
0.99015



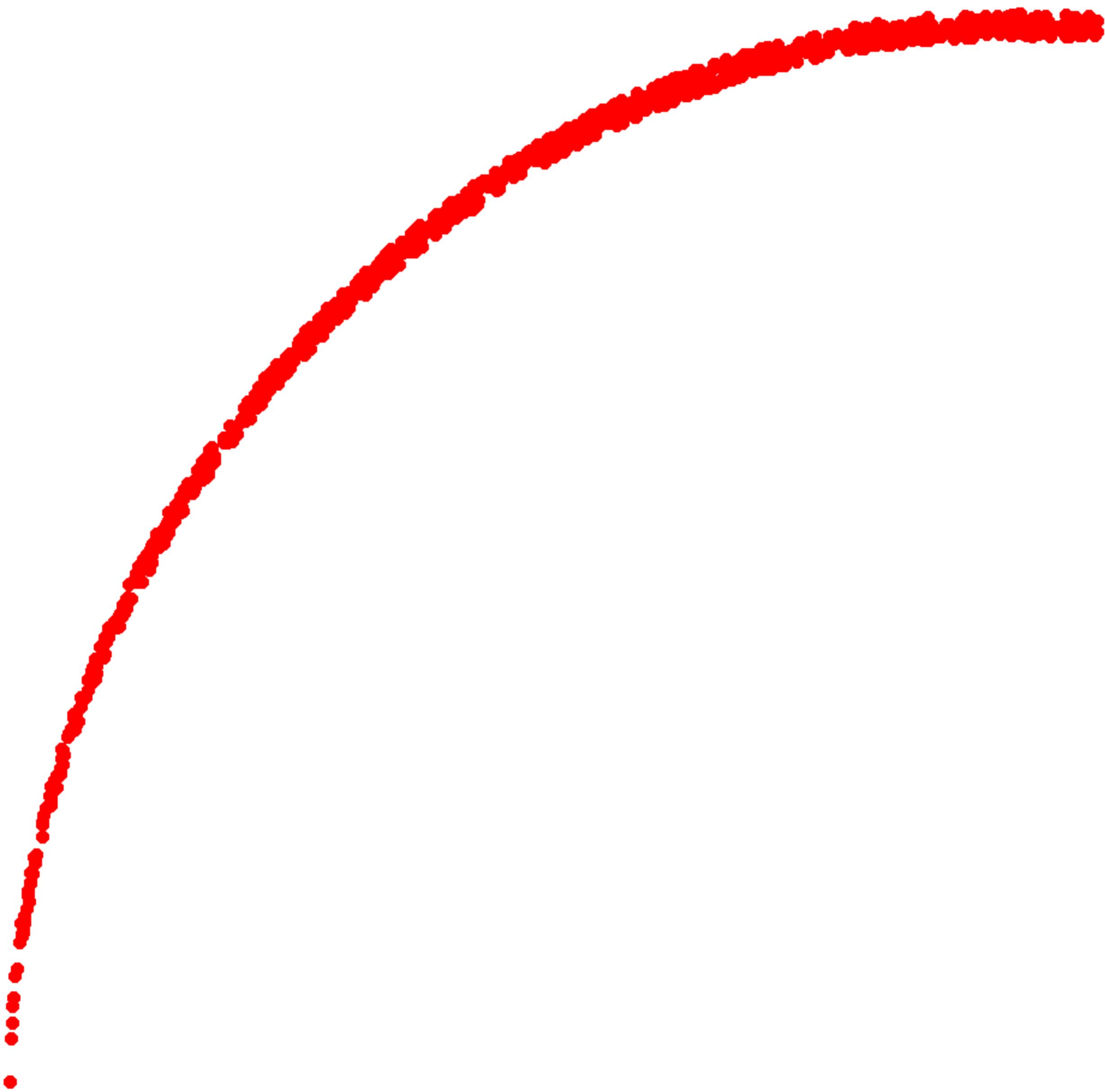
0.99163



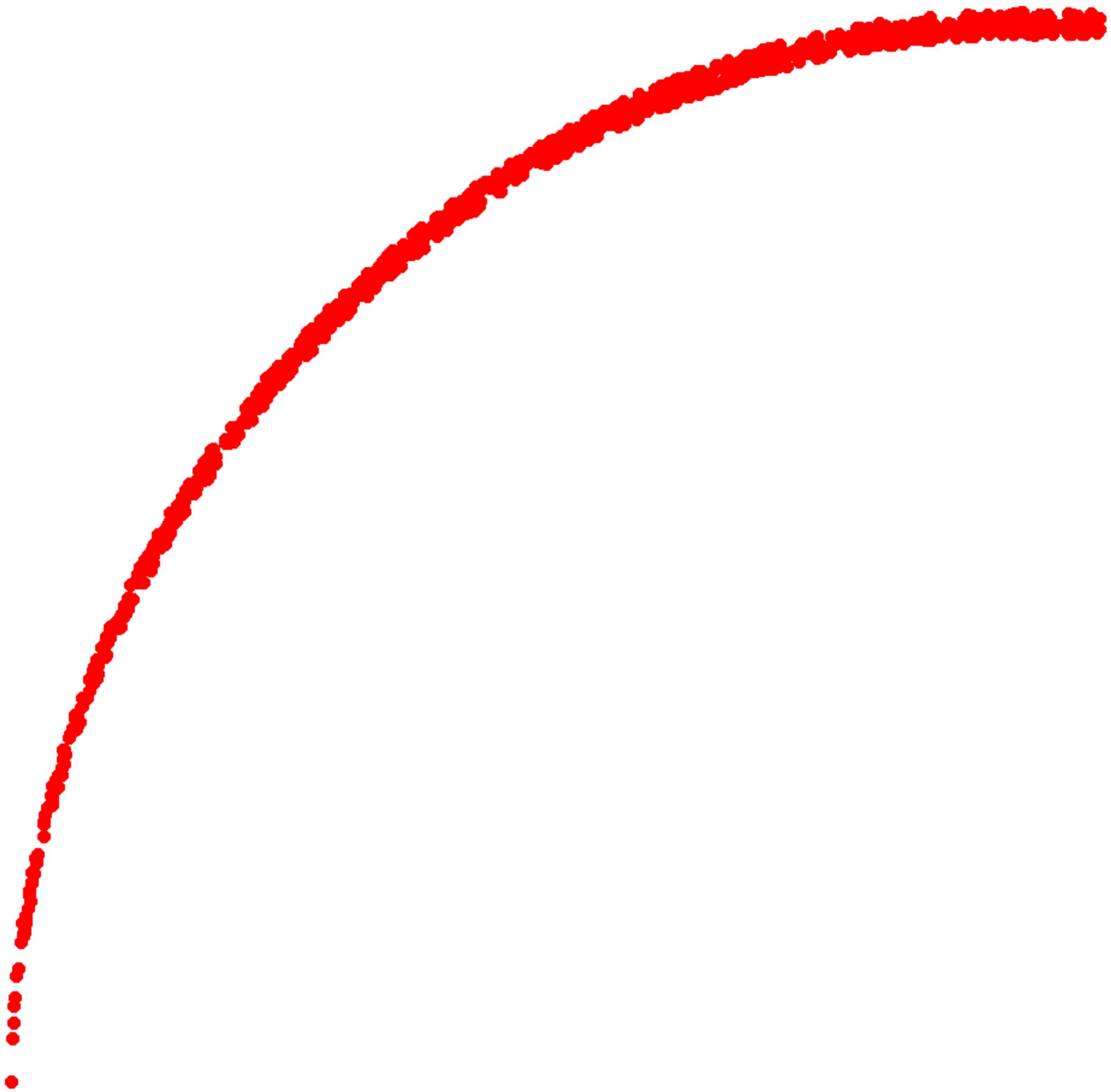
0.99276



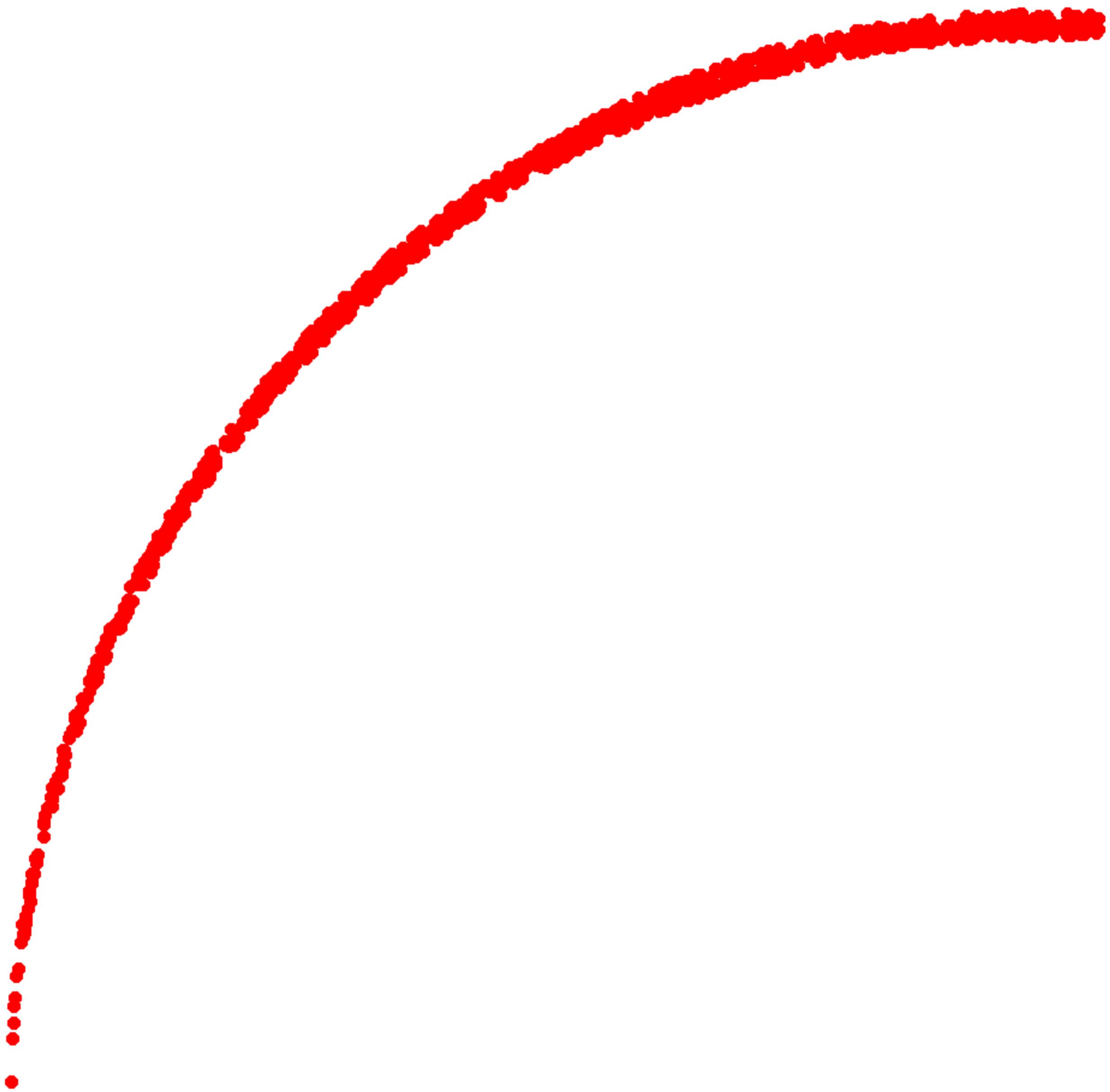
0.99376



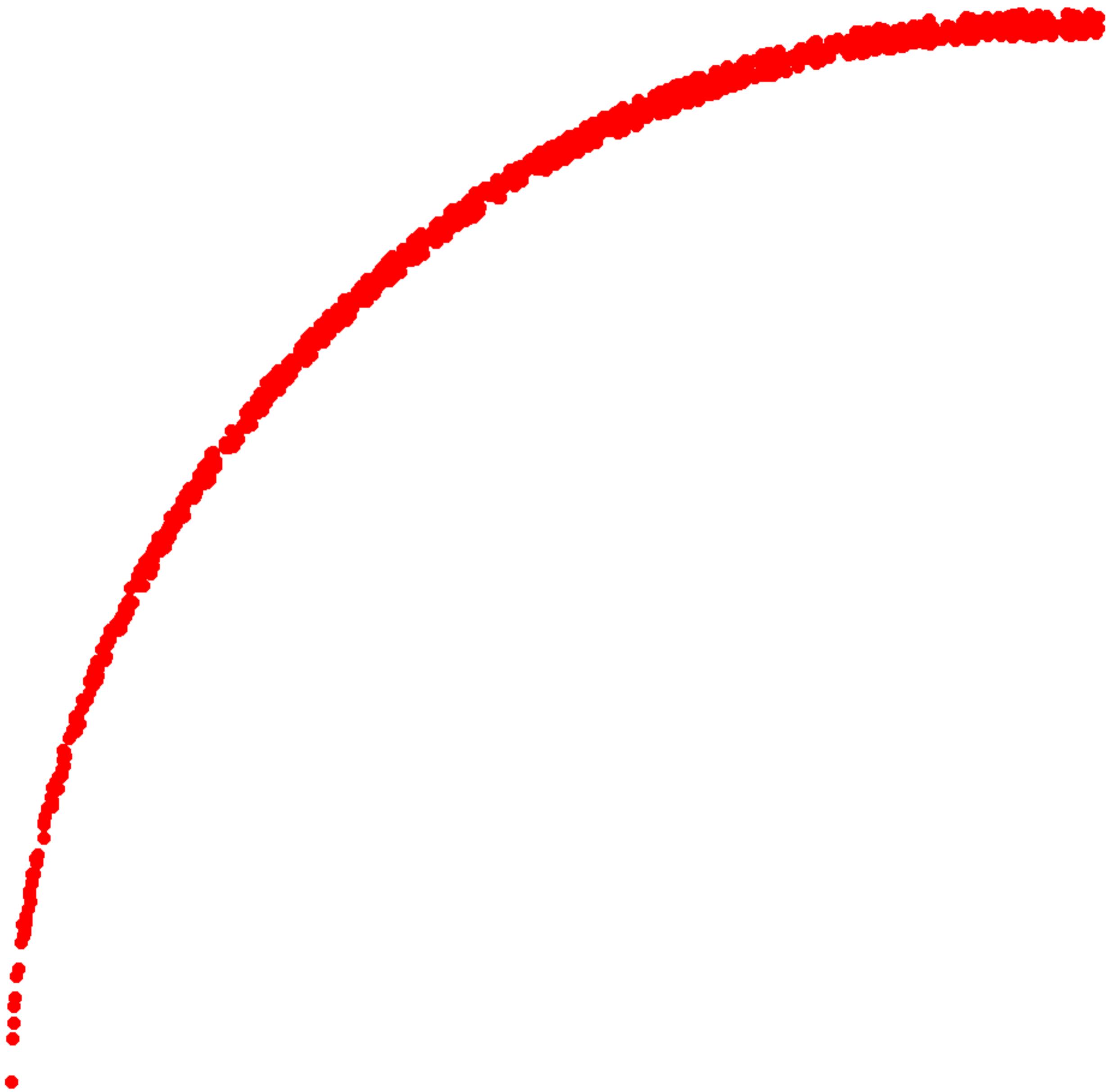
0.99454



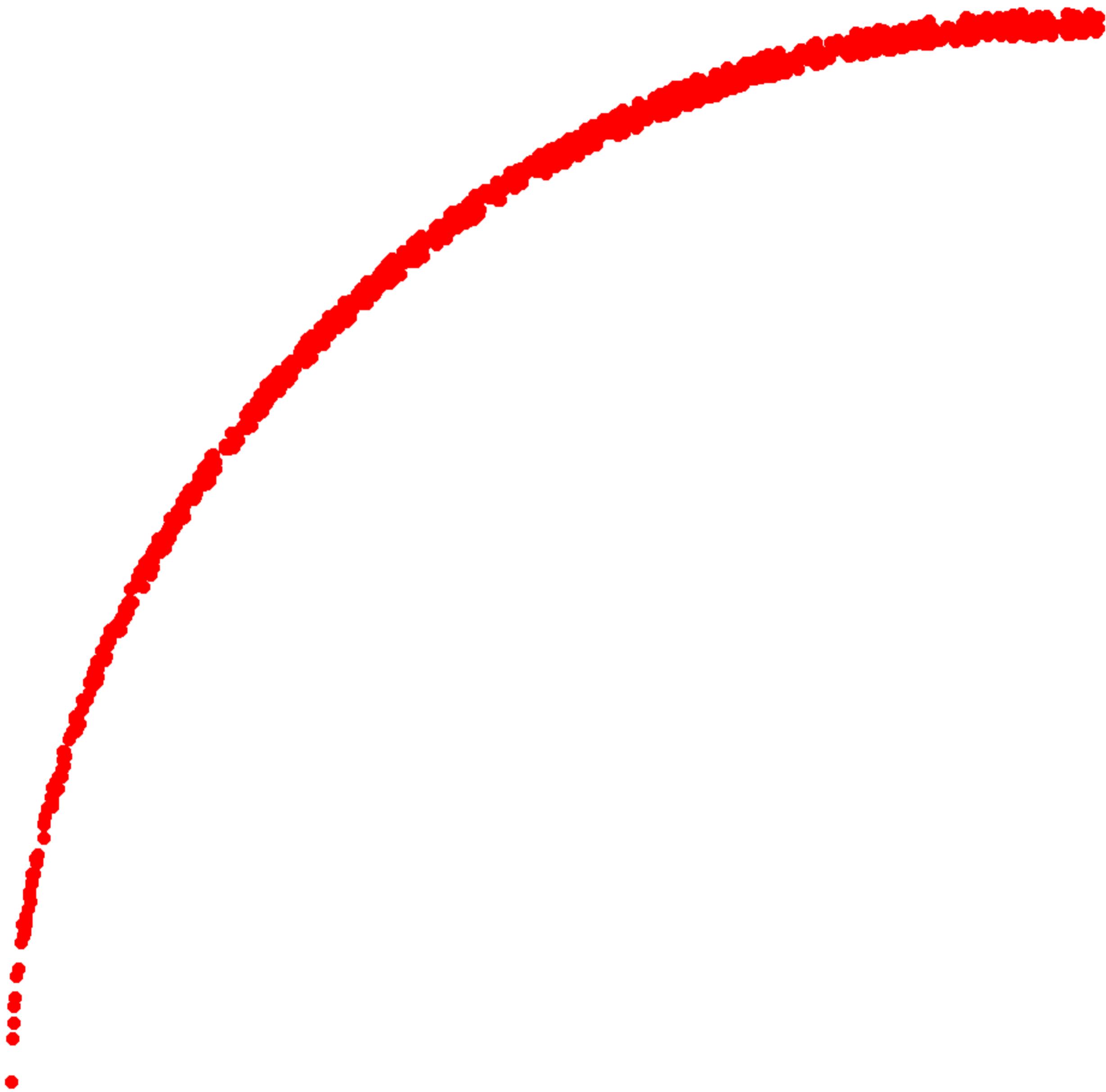
0.99515



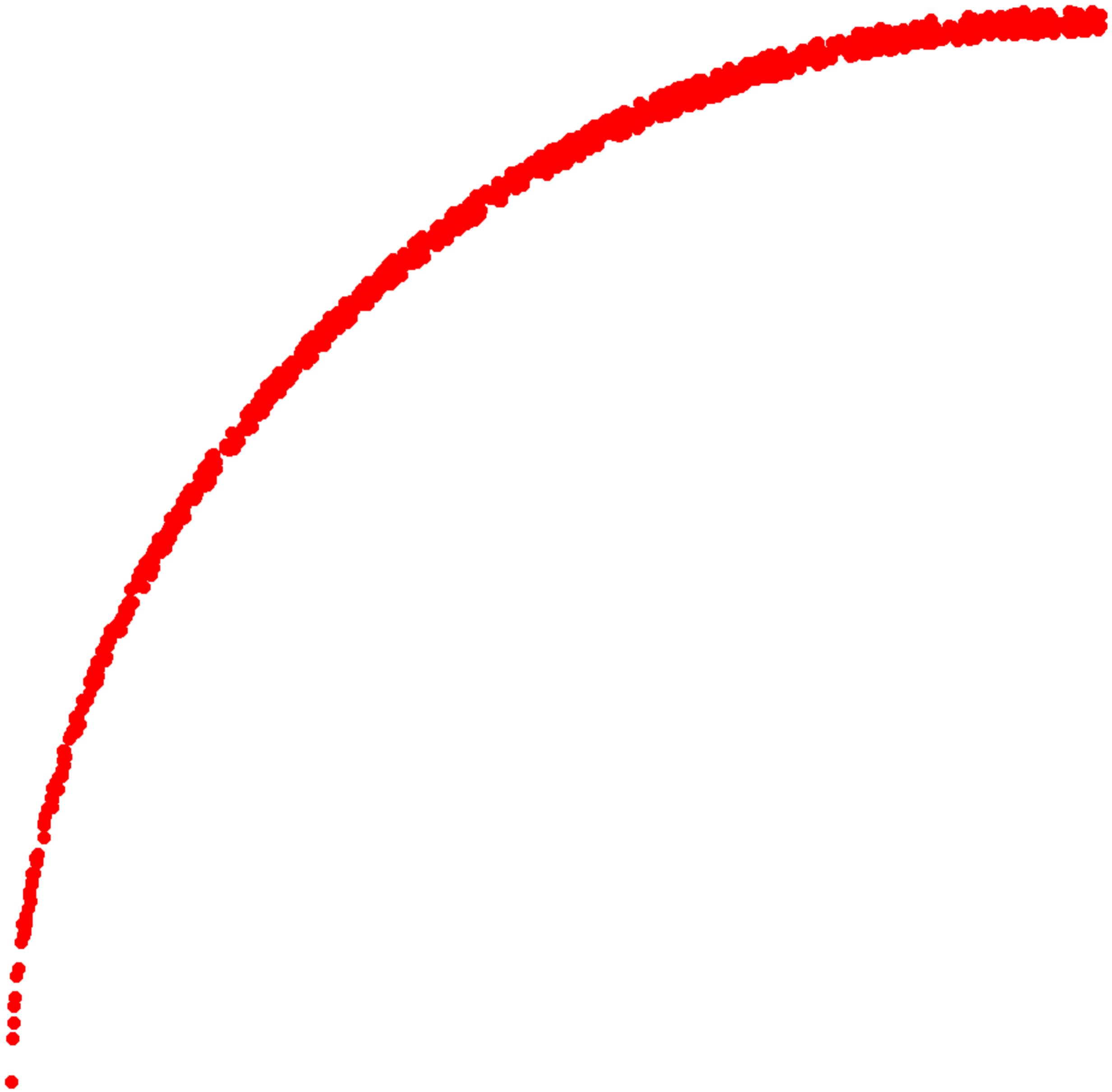
0.99566



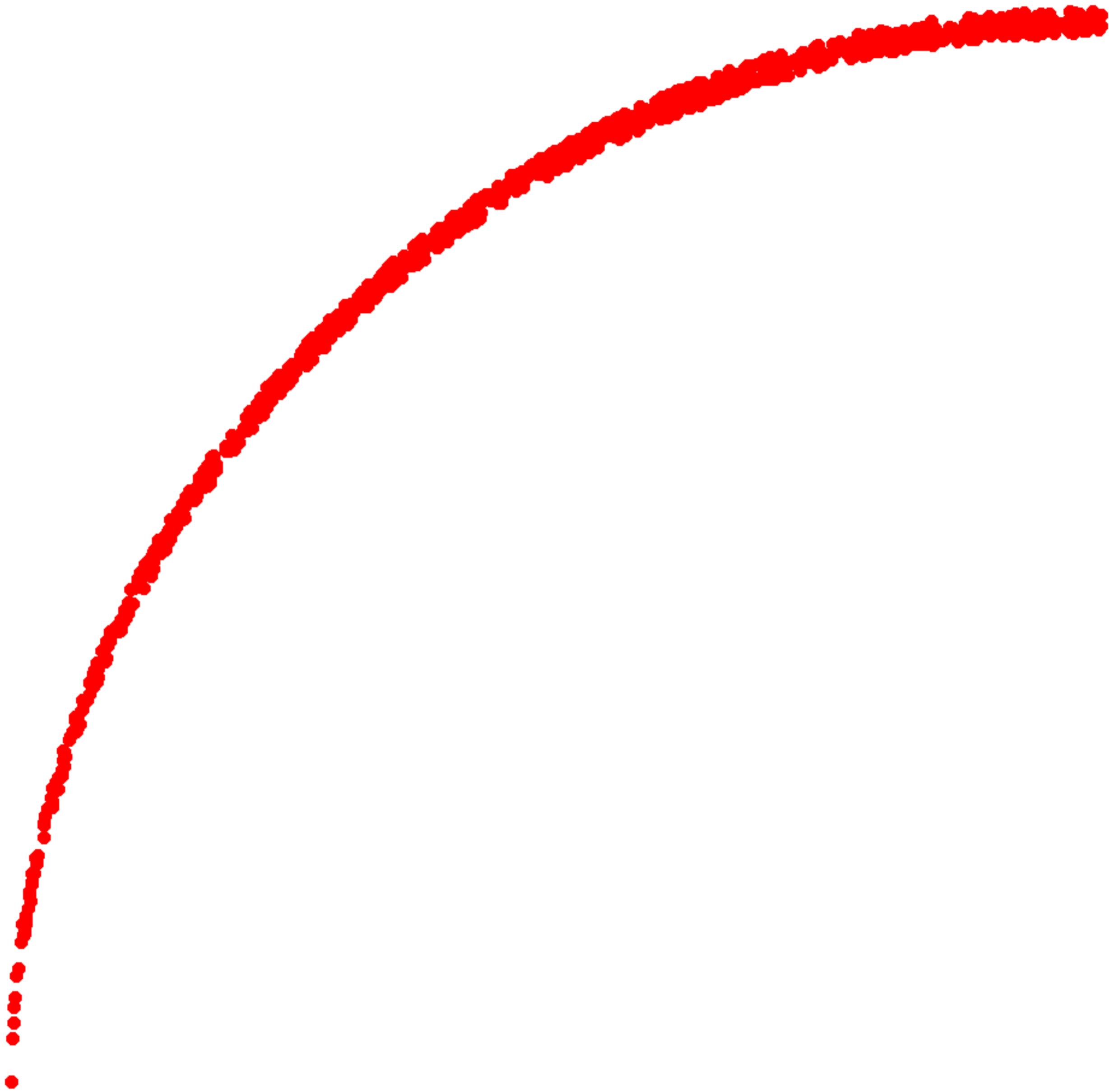
0.99612



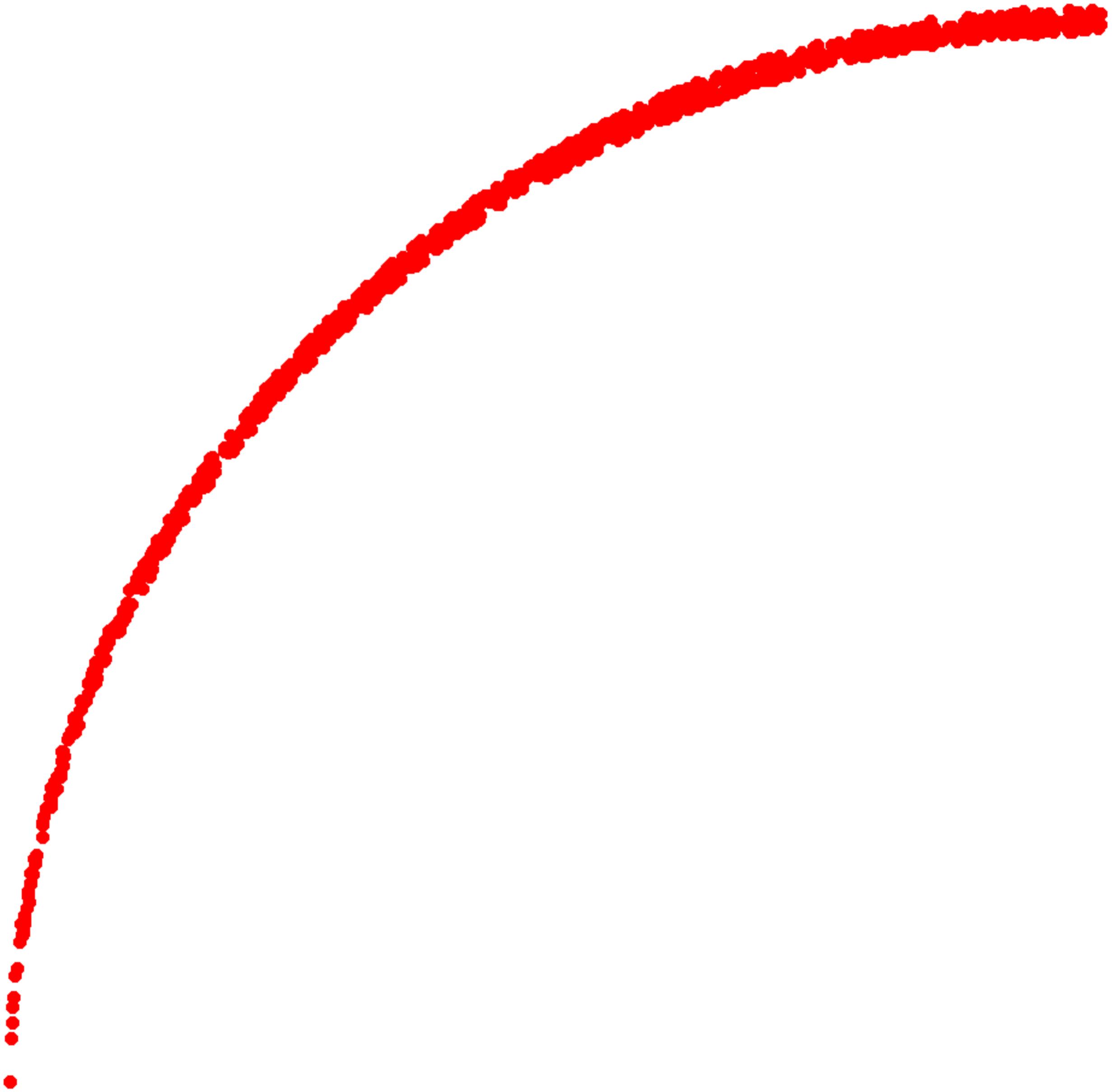
0.99651



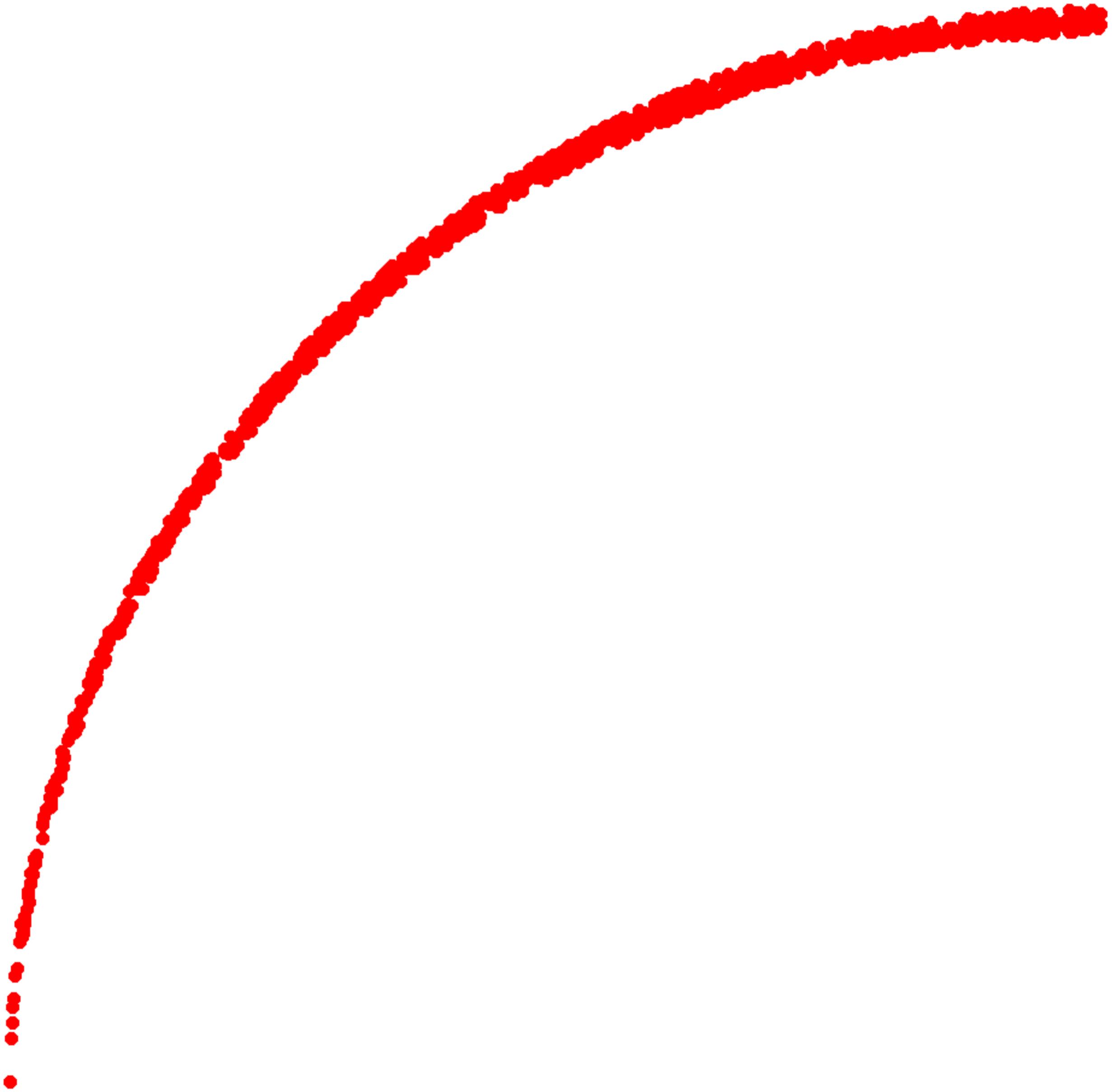
0.99681



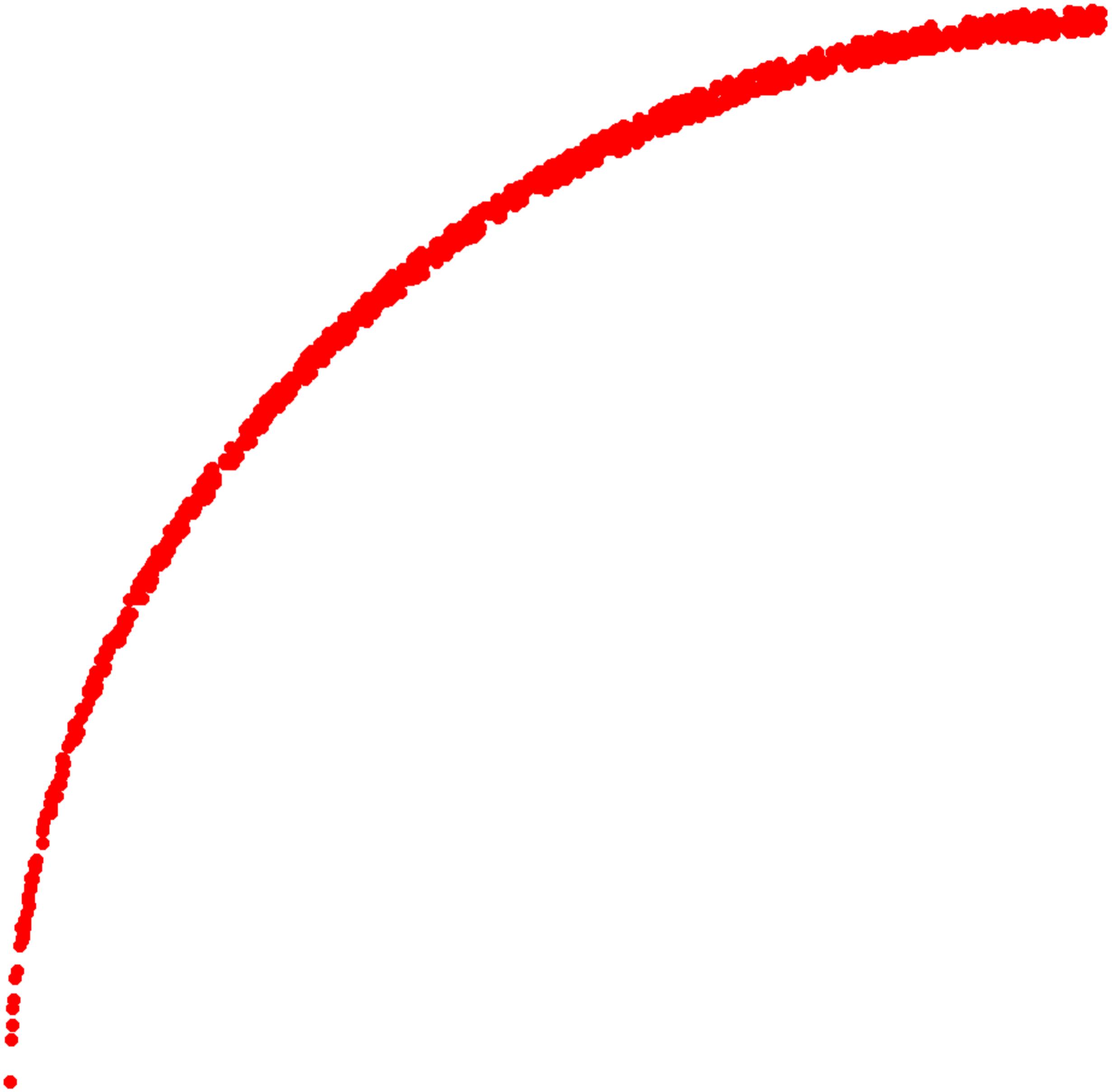
0.99707



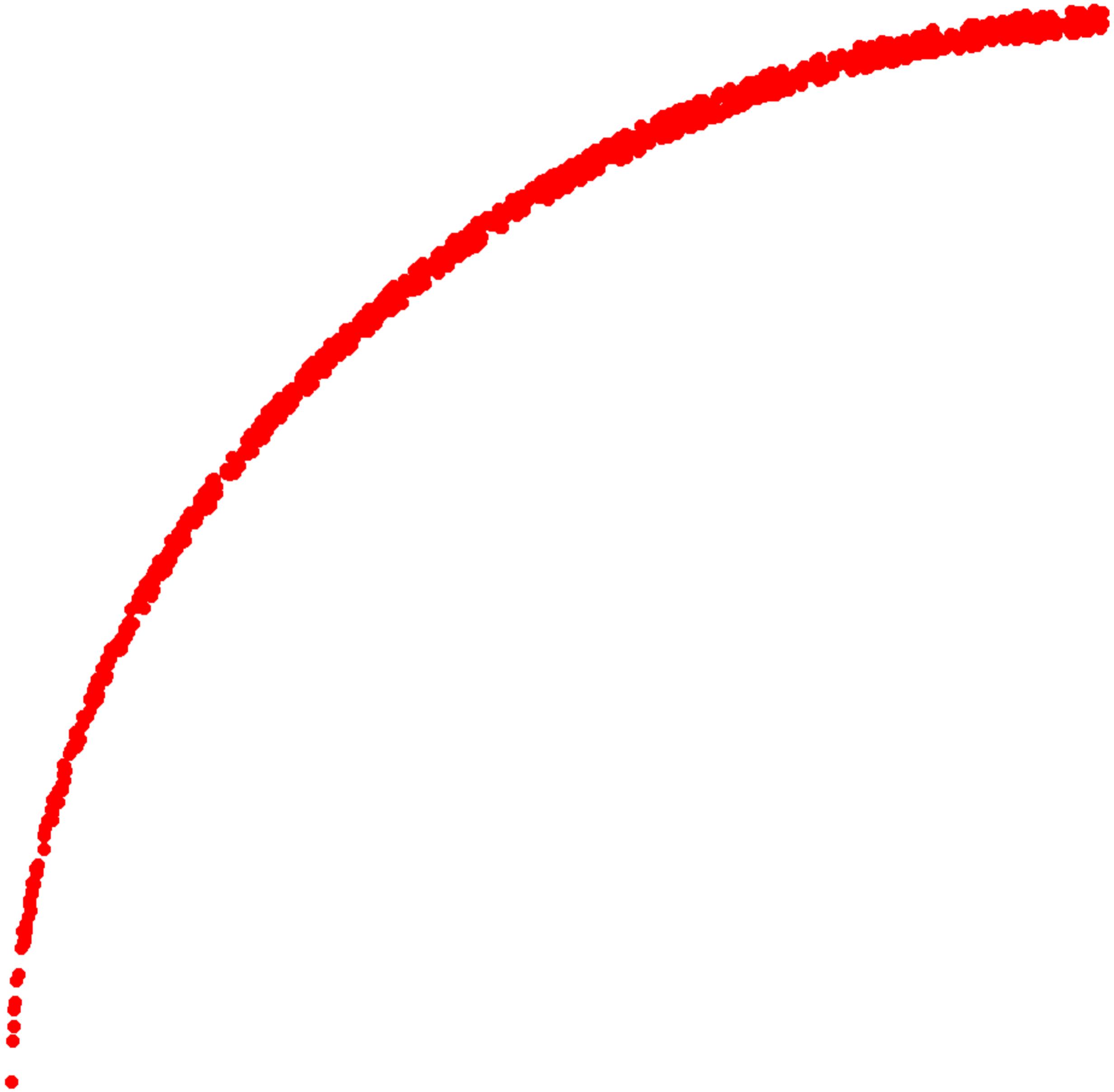
0.99725



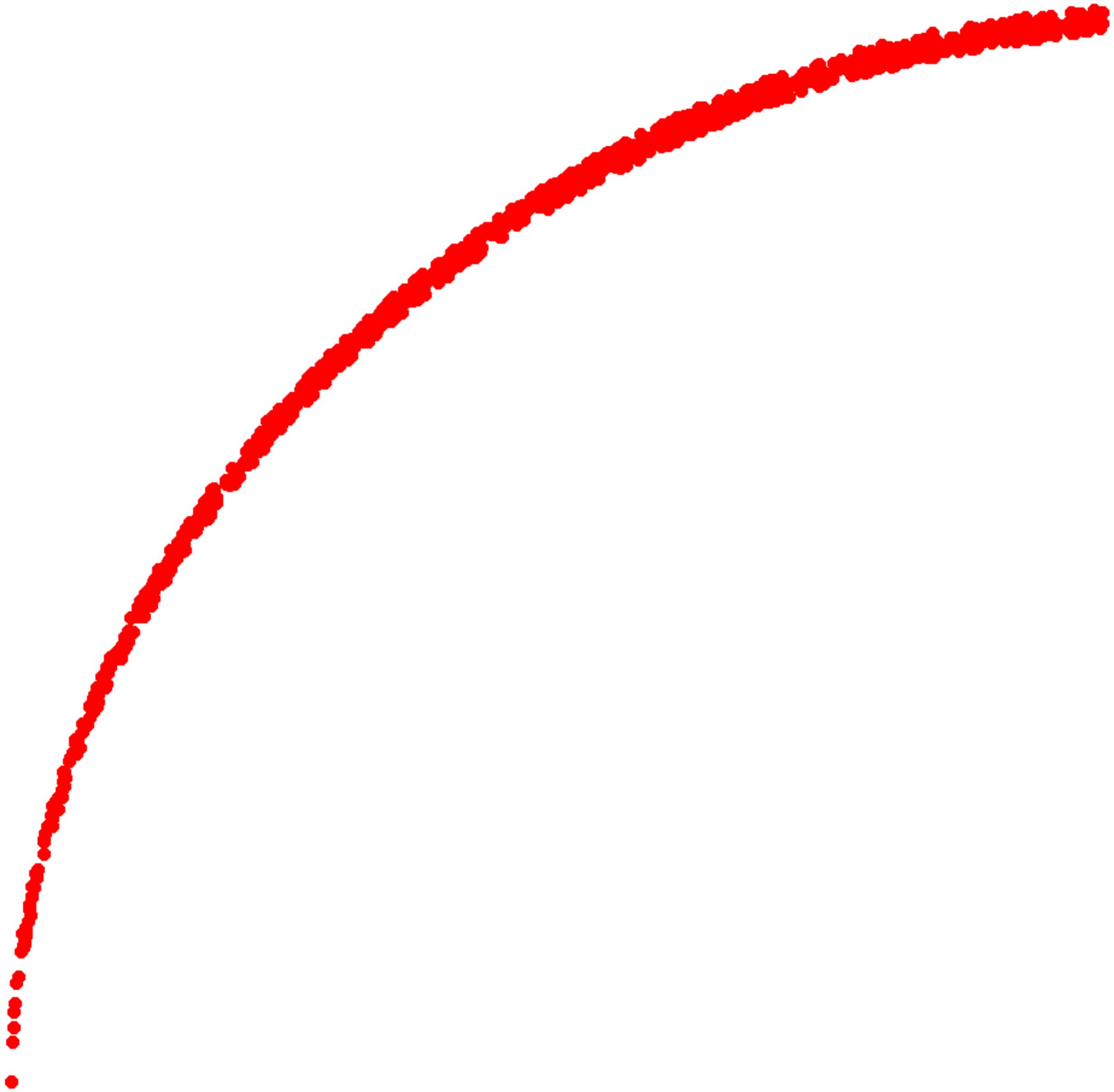
0.99774



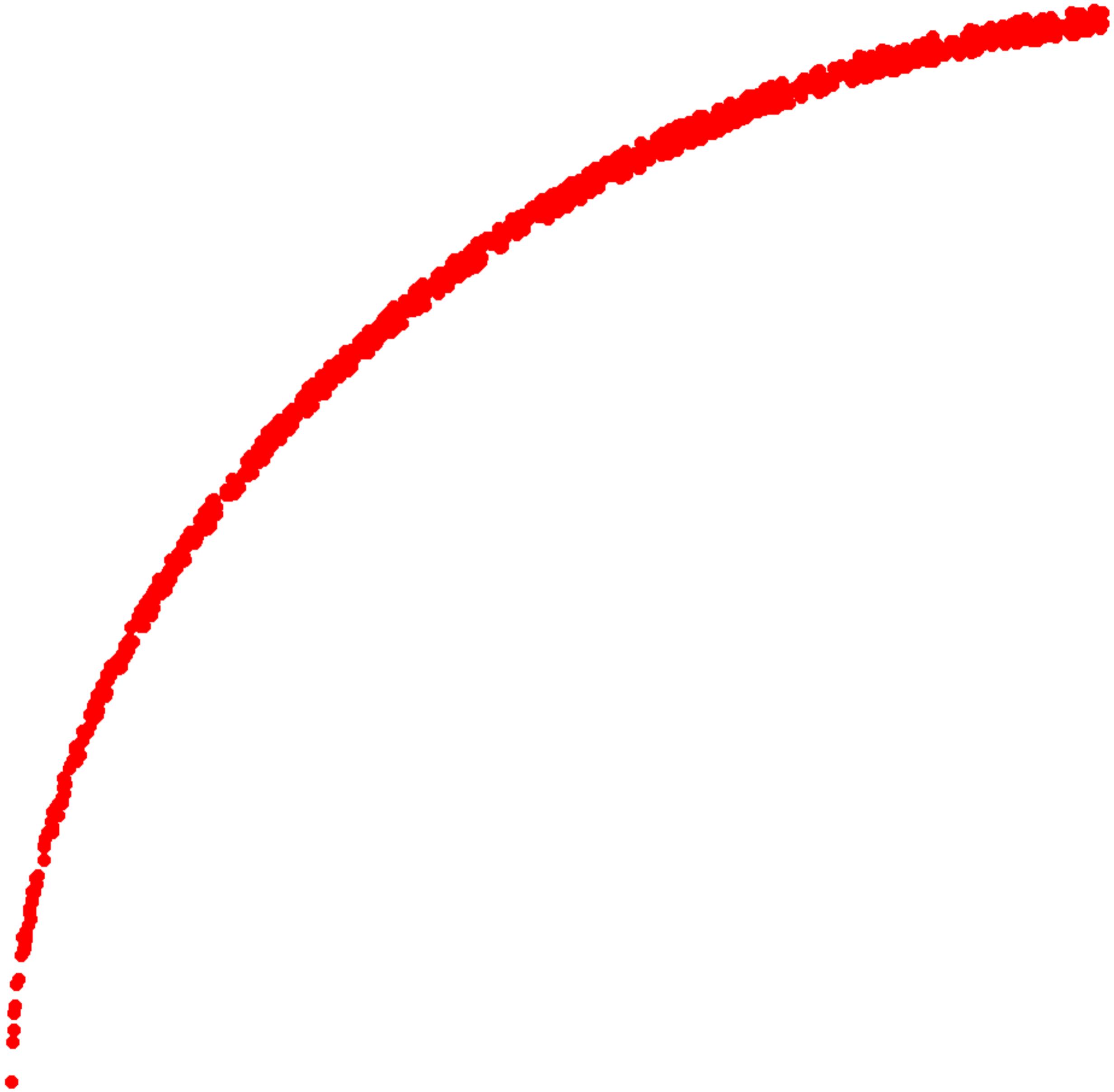
0.99806



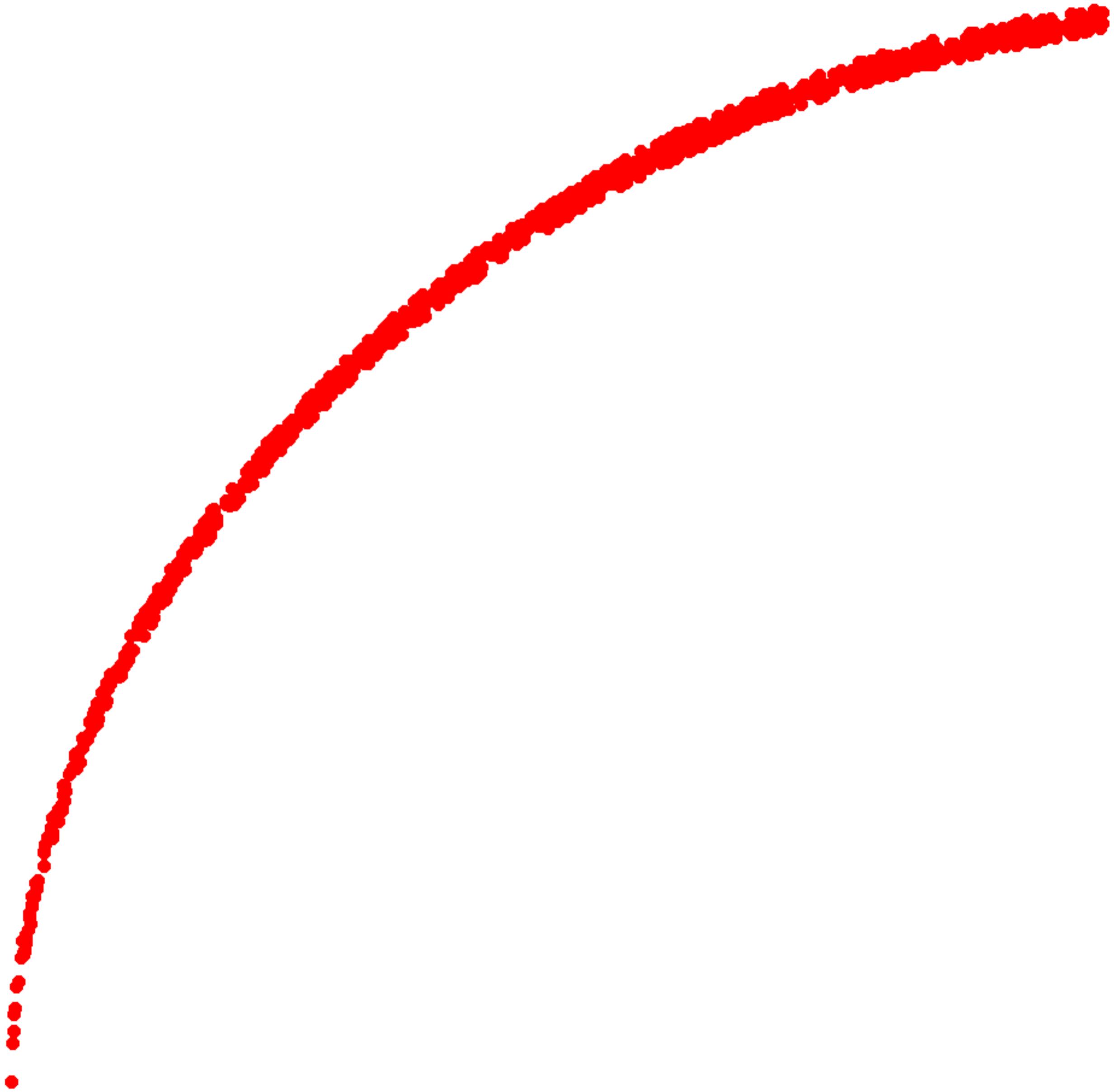
0.99831



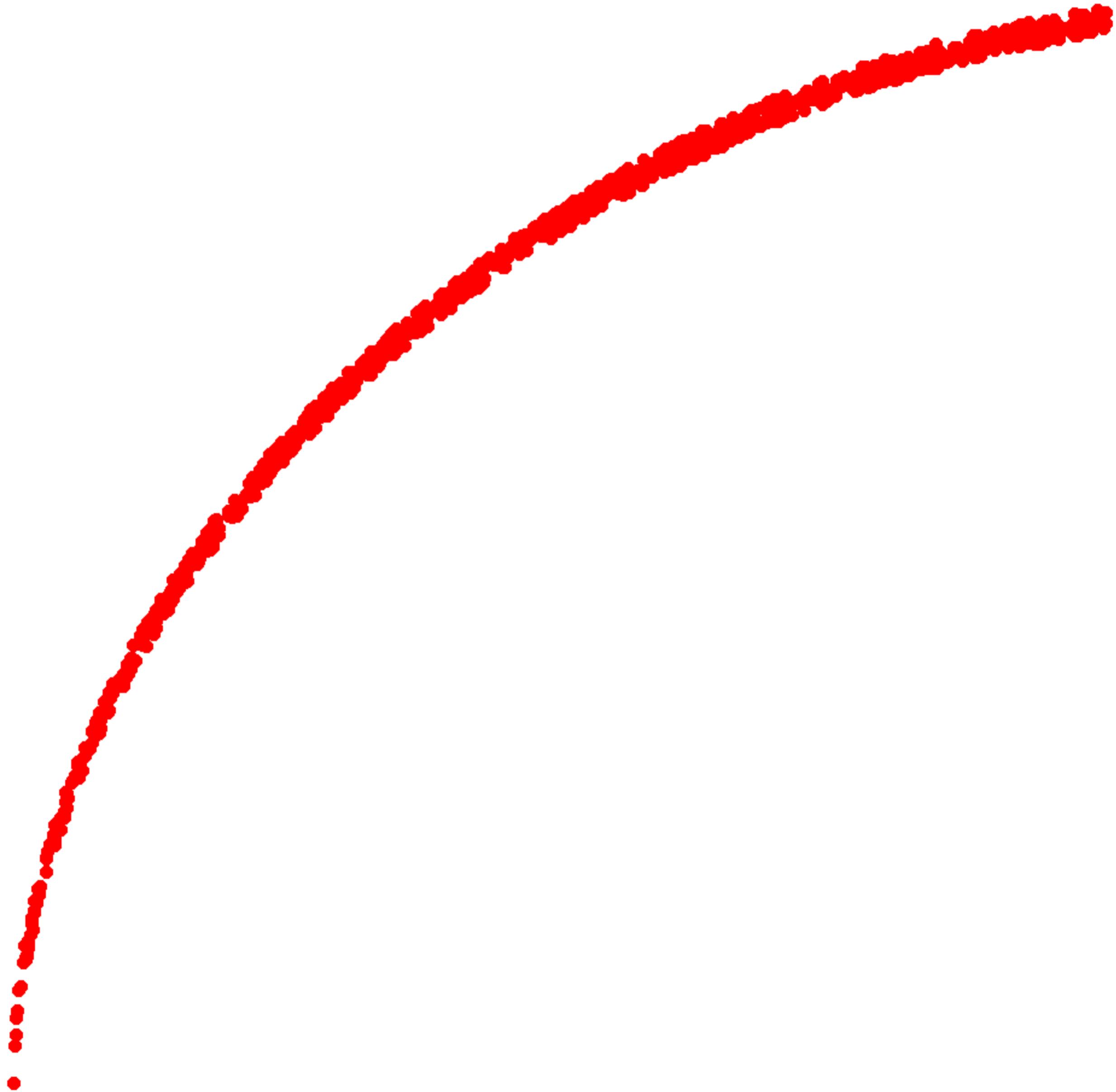
0.99853



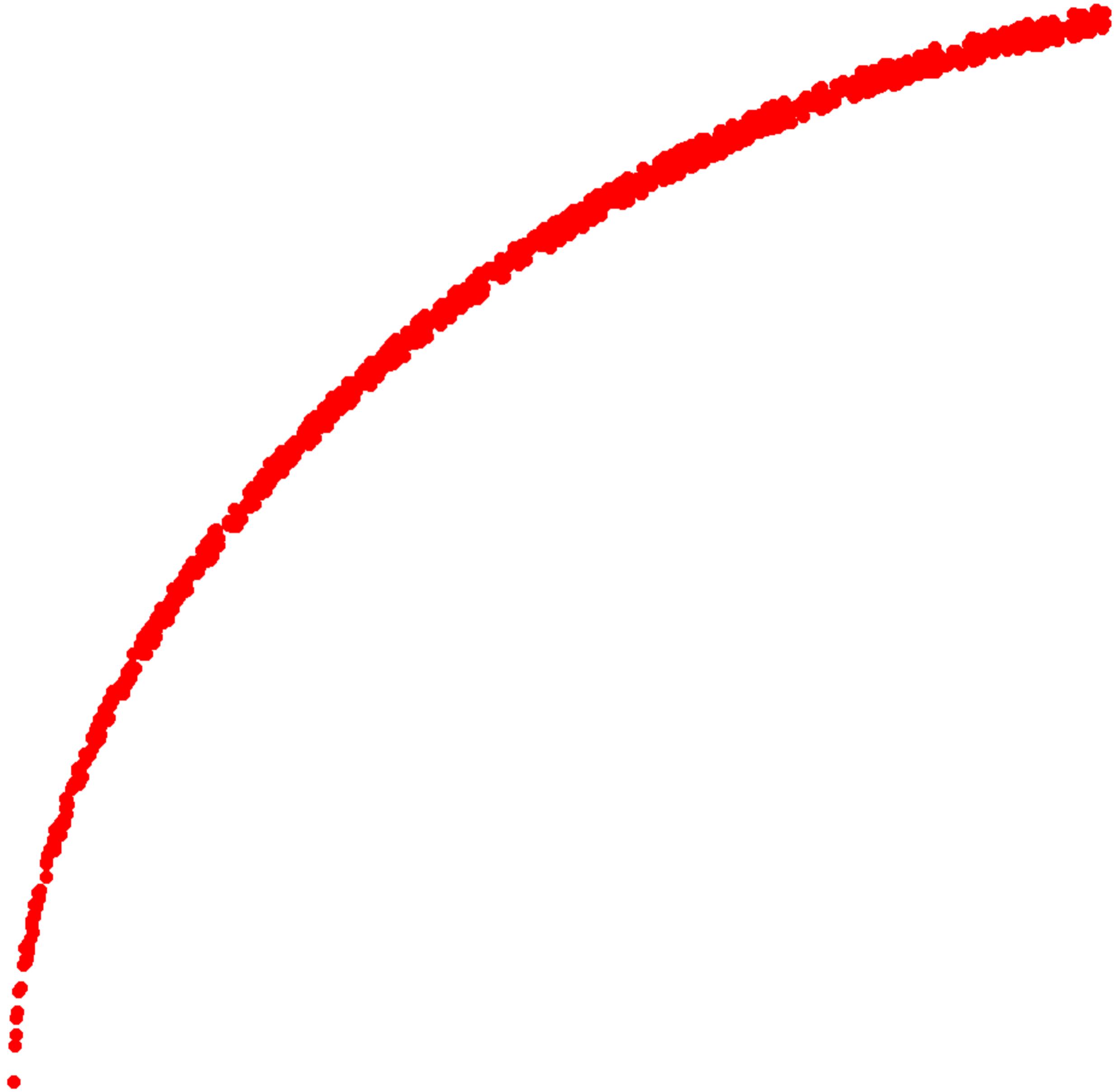
0.99866



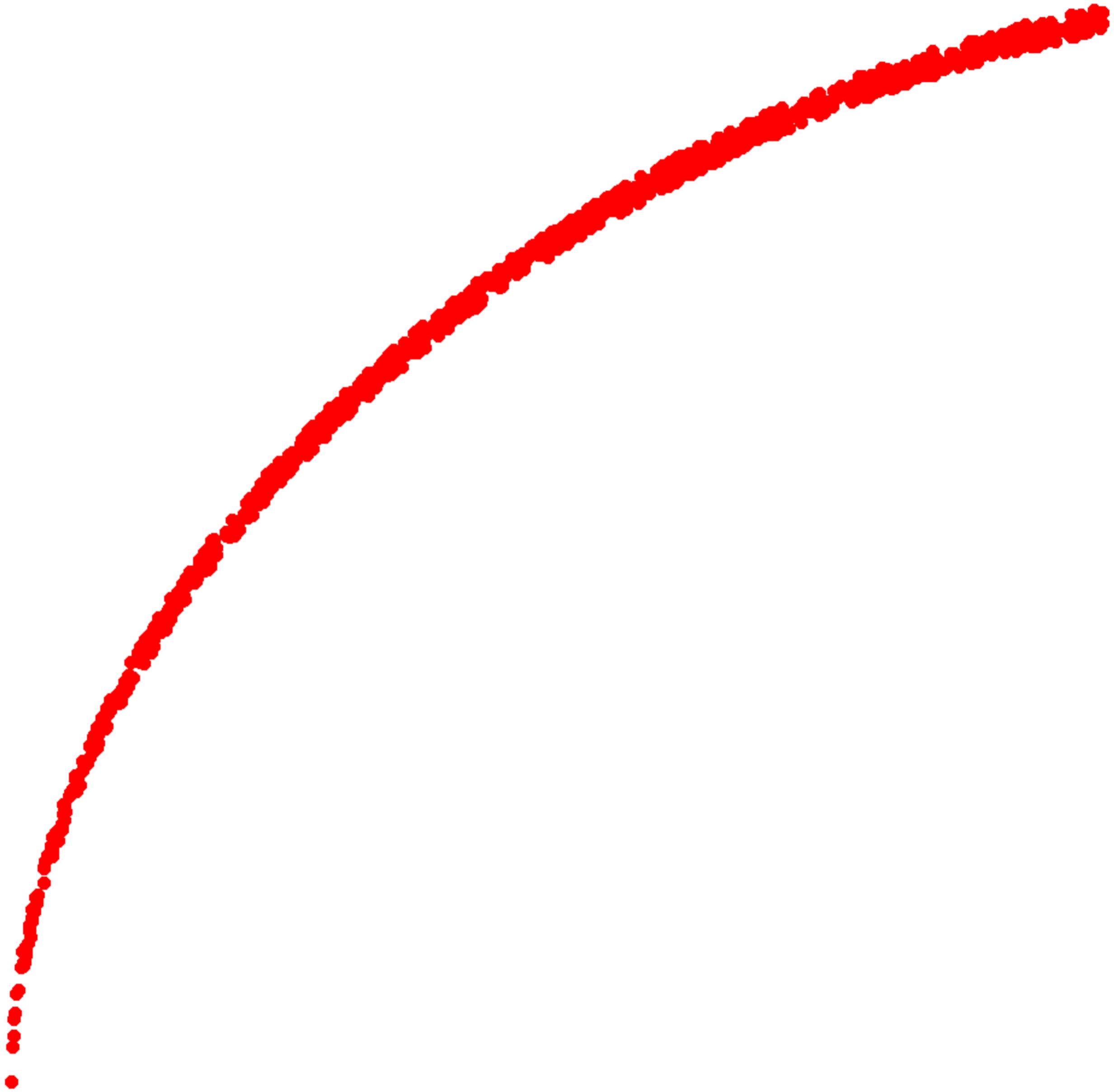
0.99879



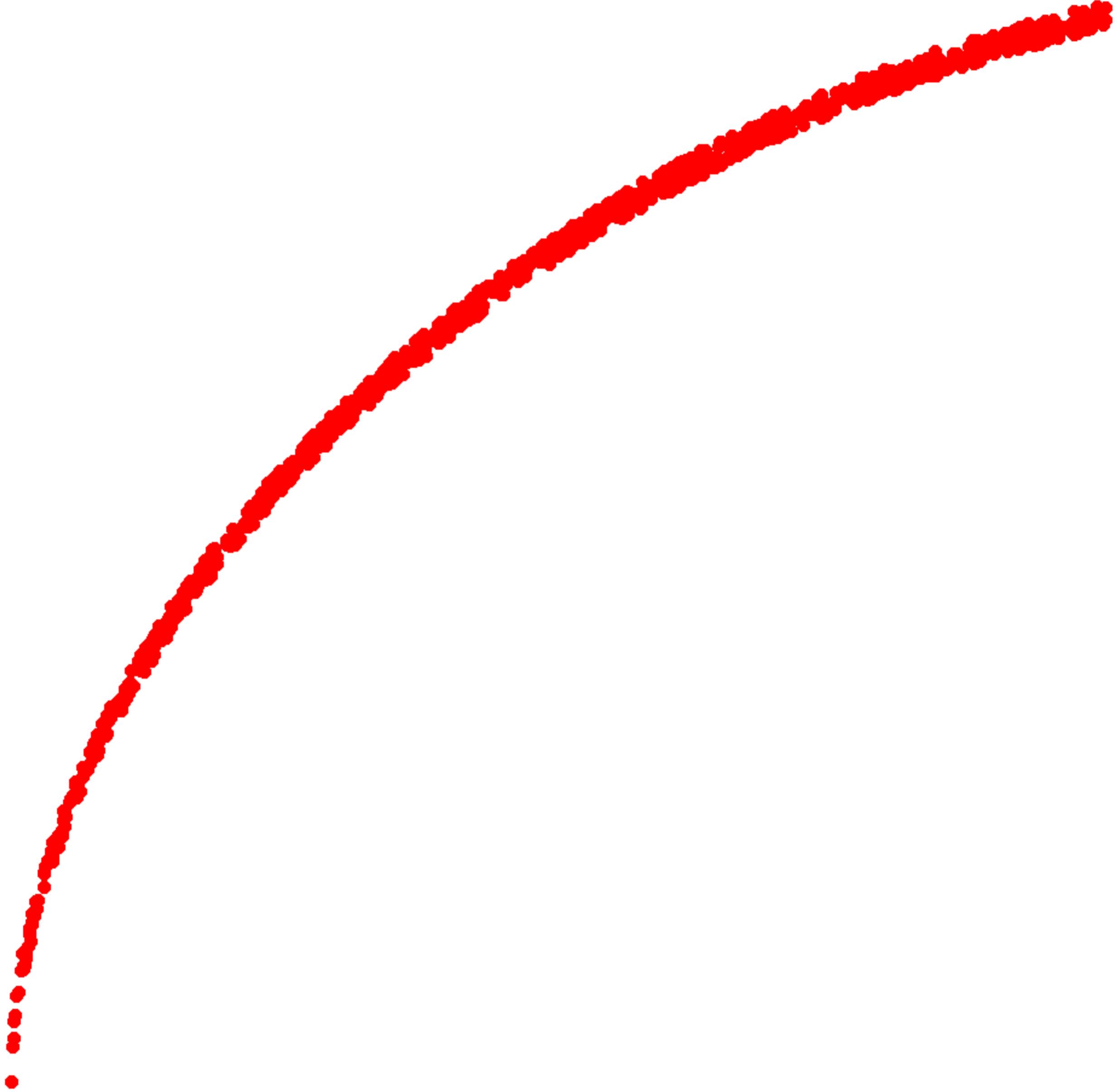
0.9989



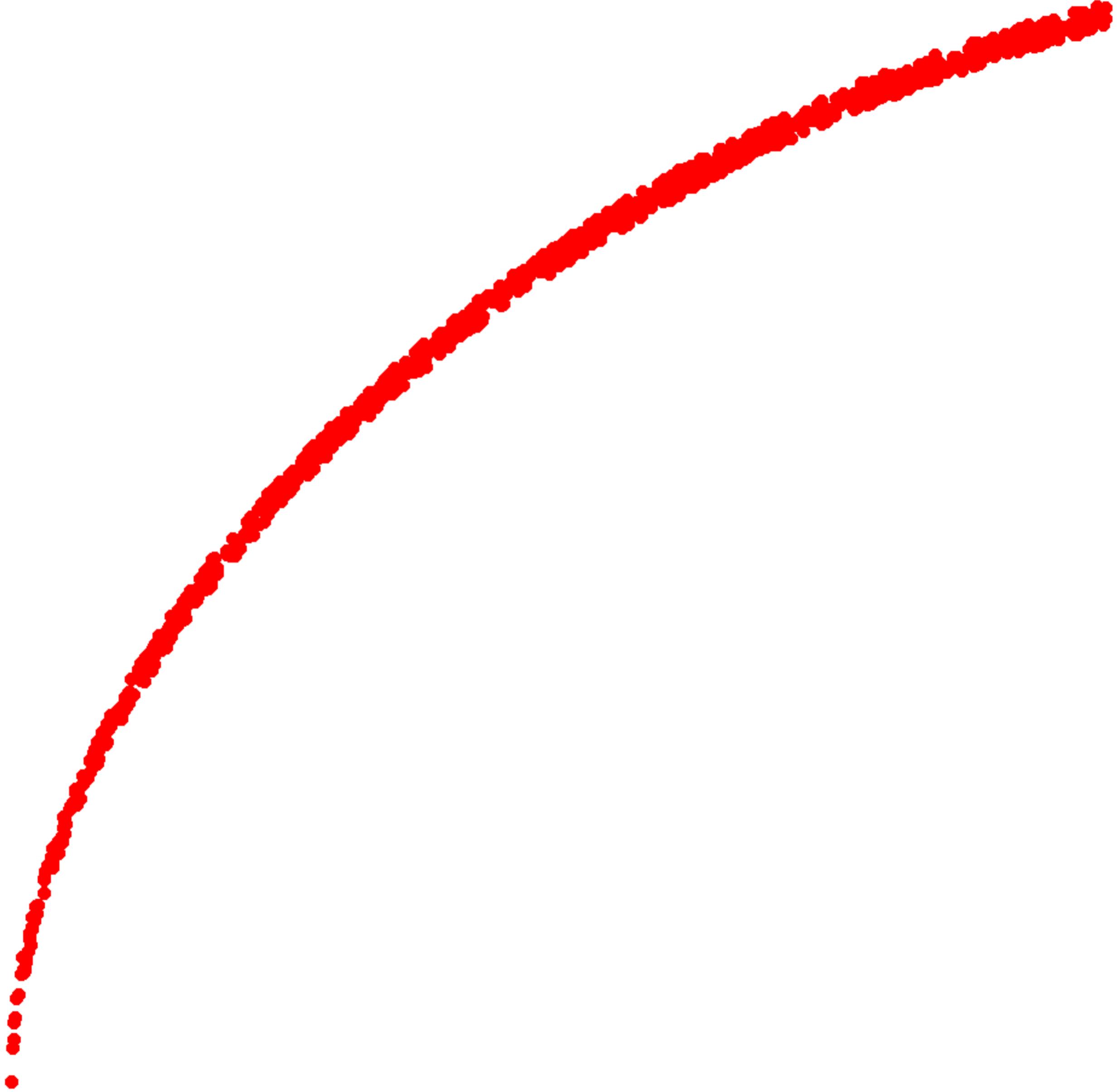
0.99899



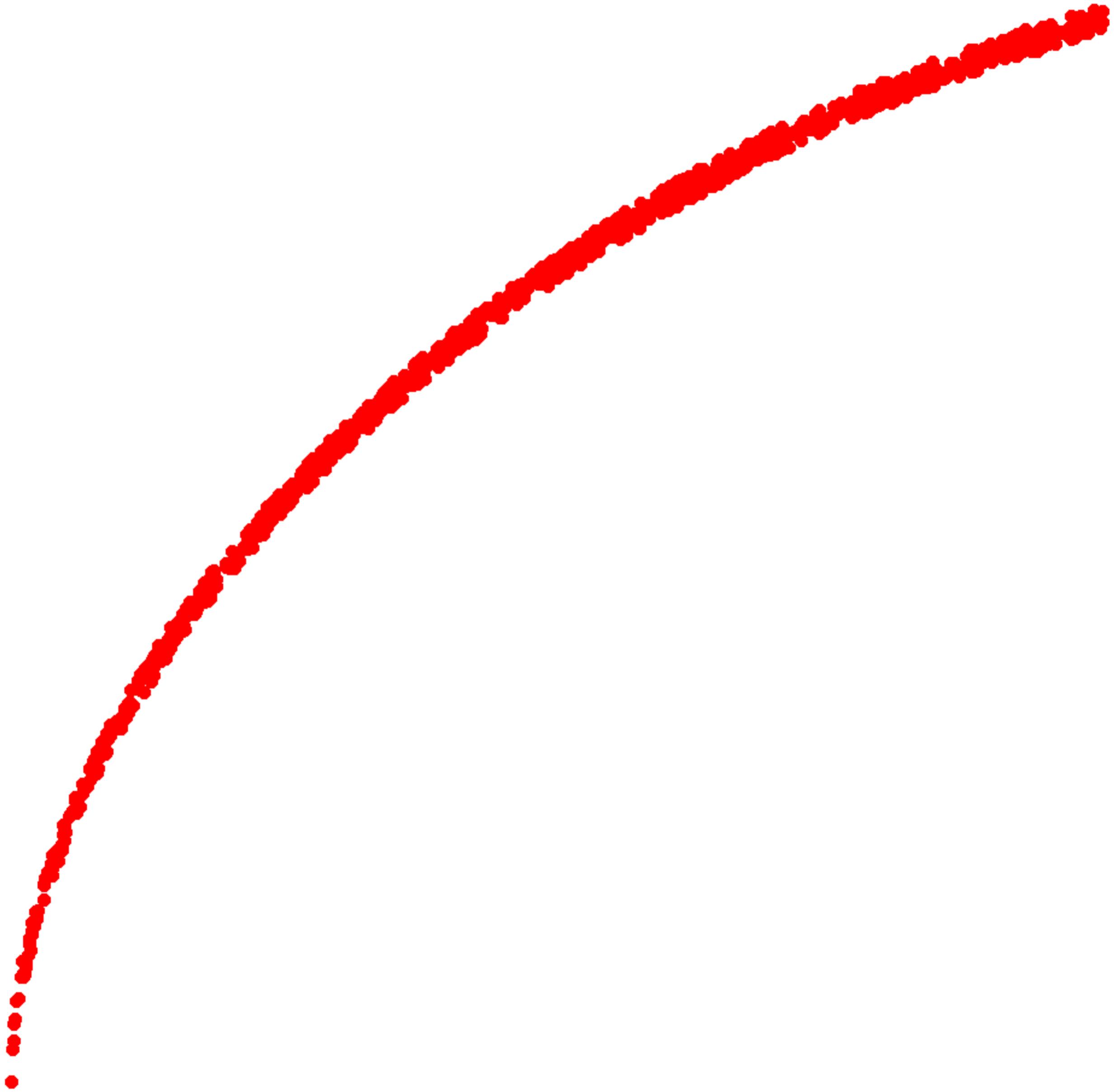
0.99906



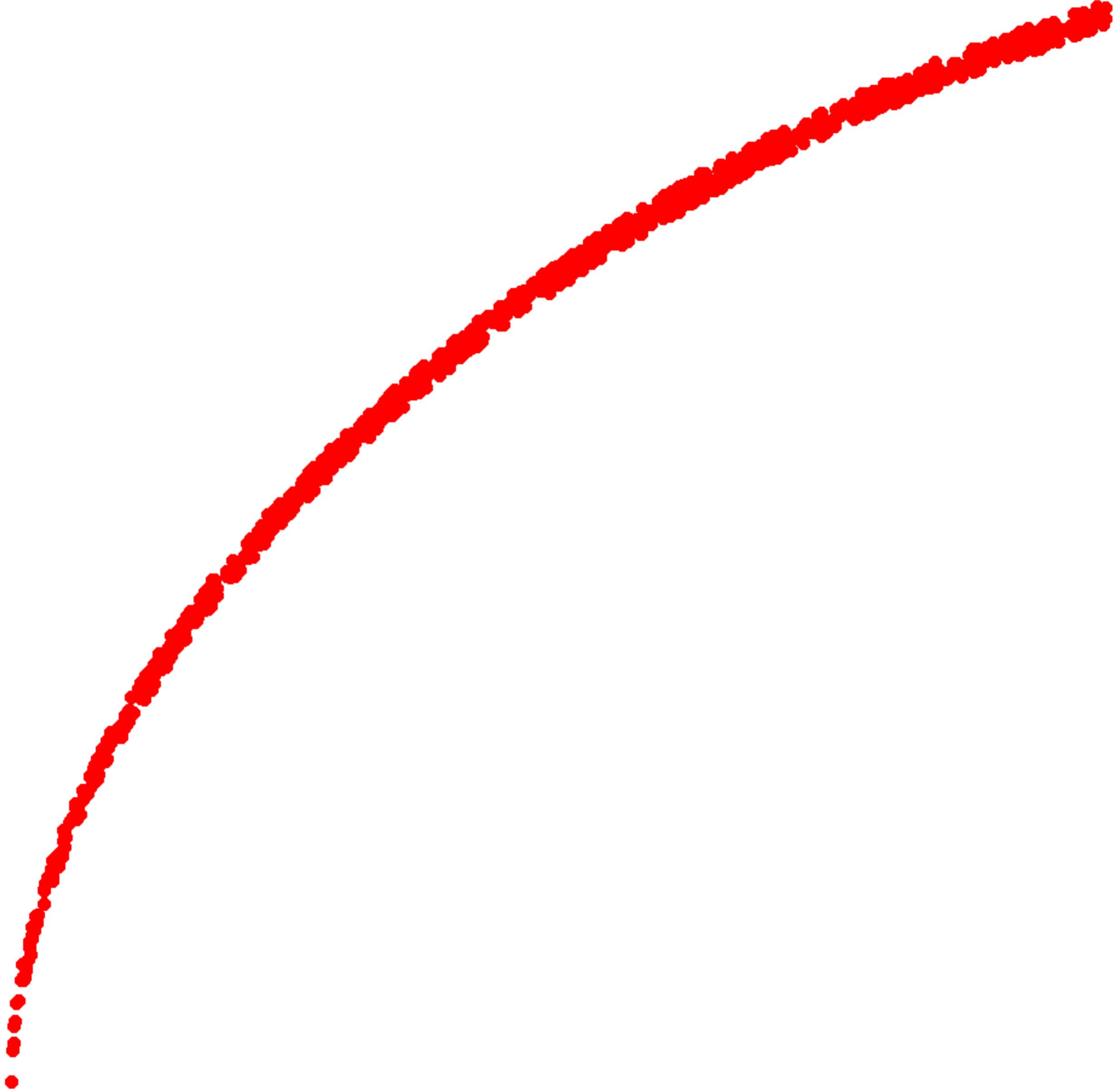
0.99912



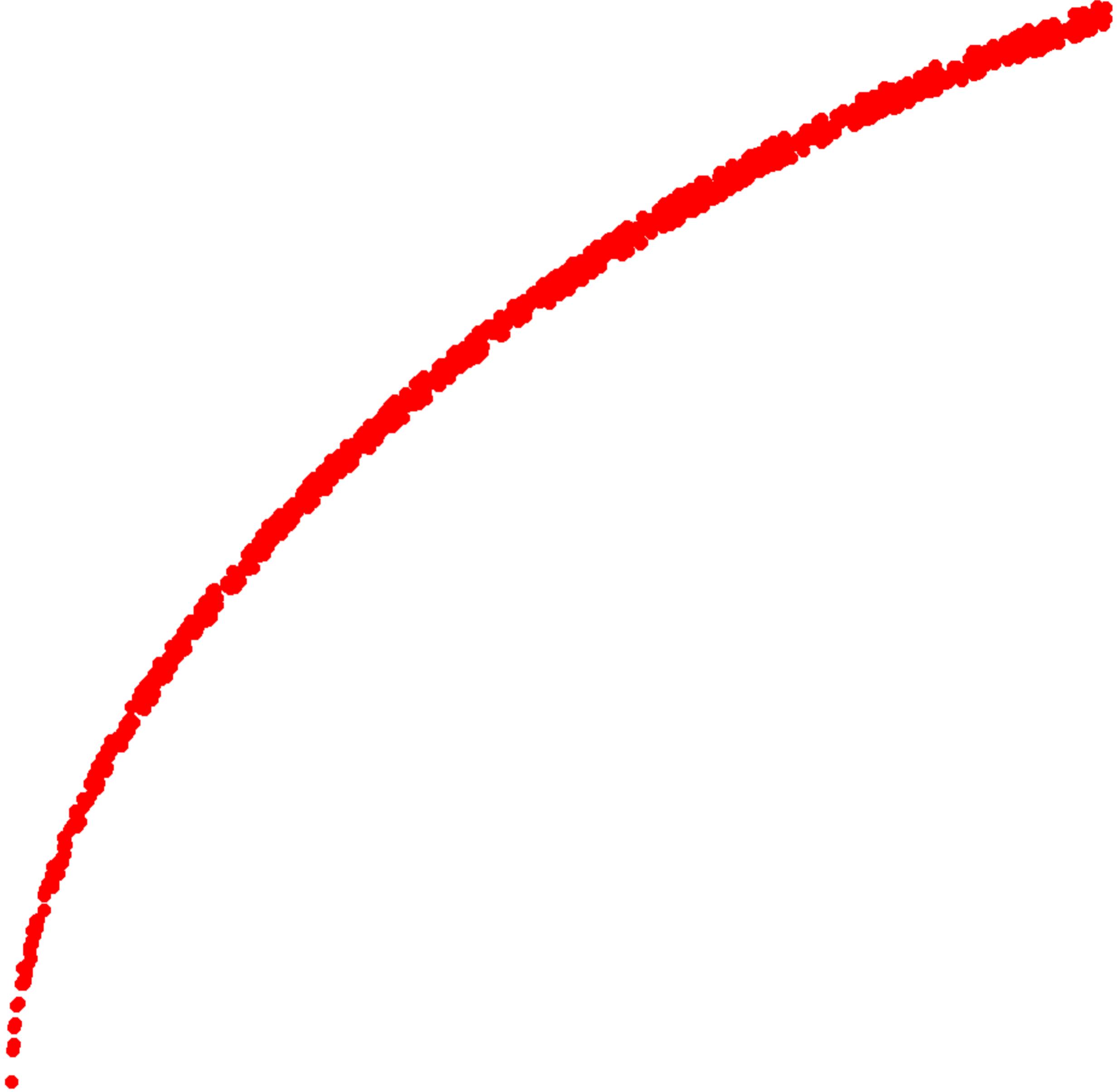
0.99918



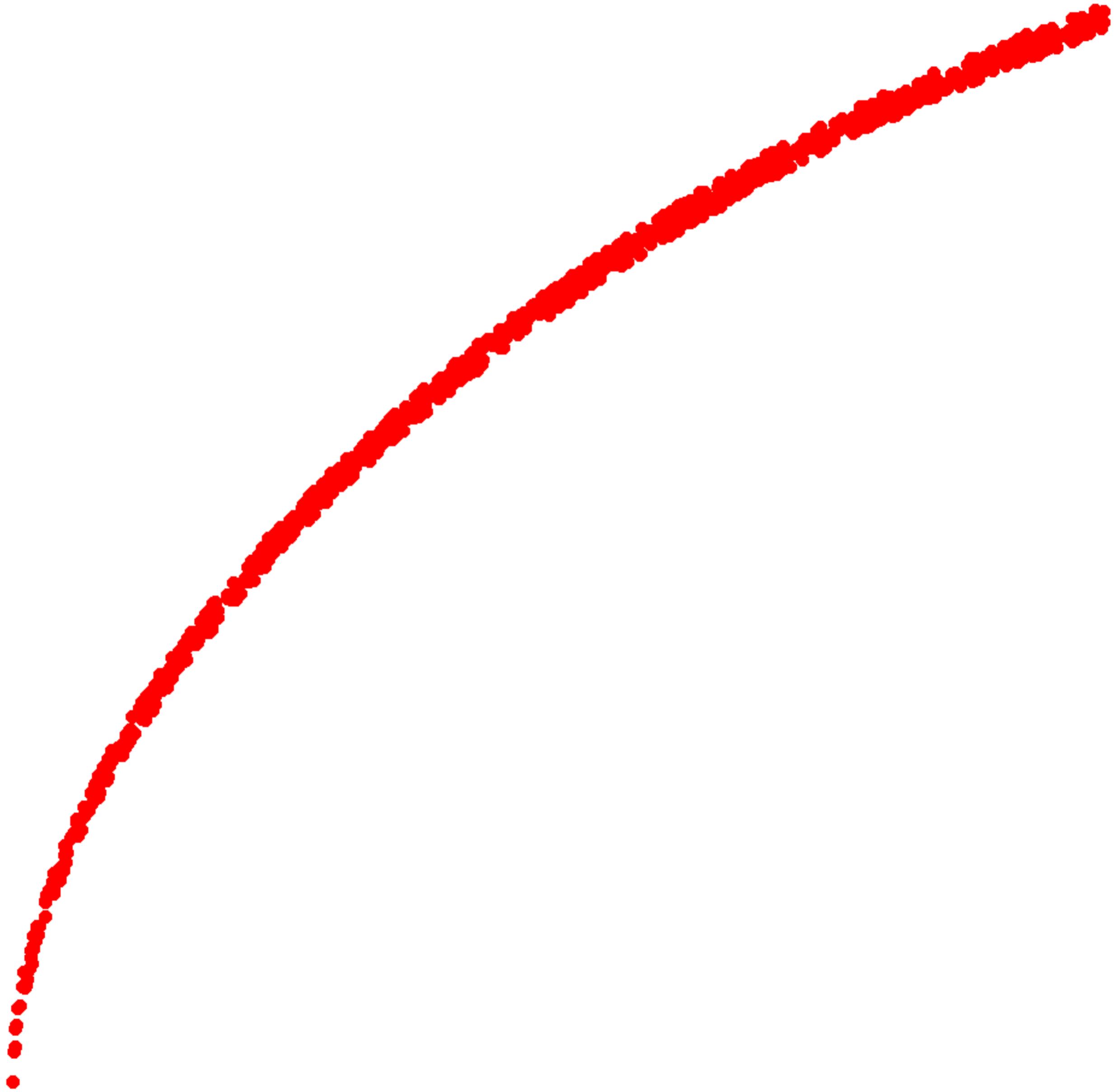
0.99922



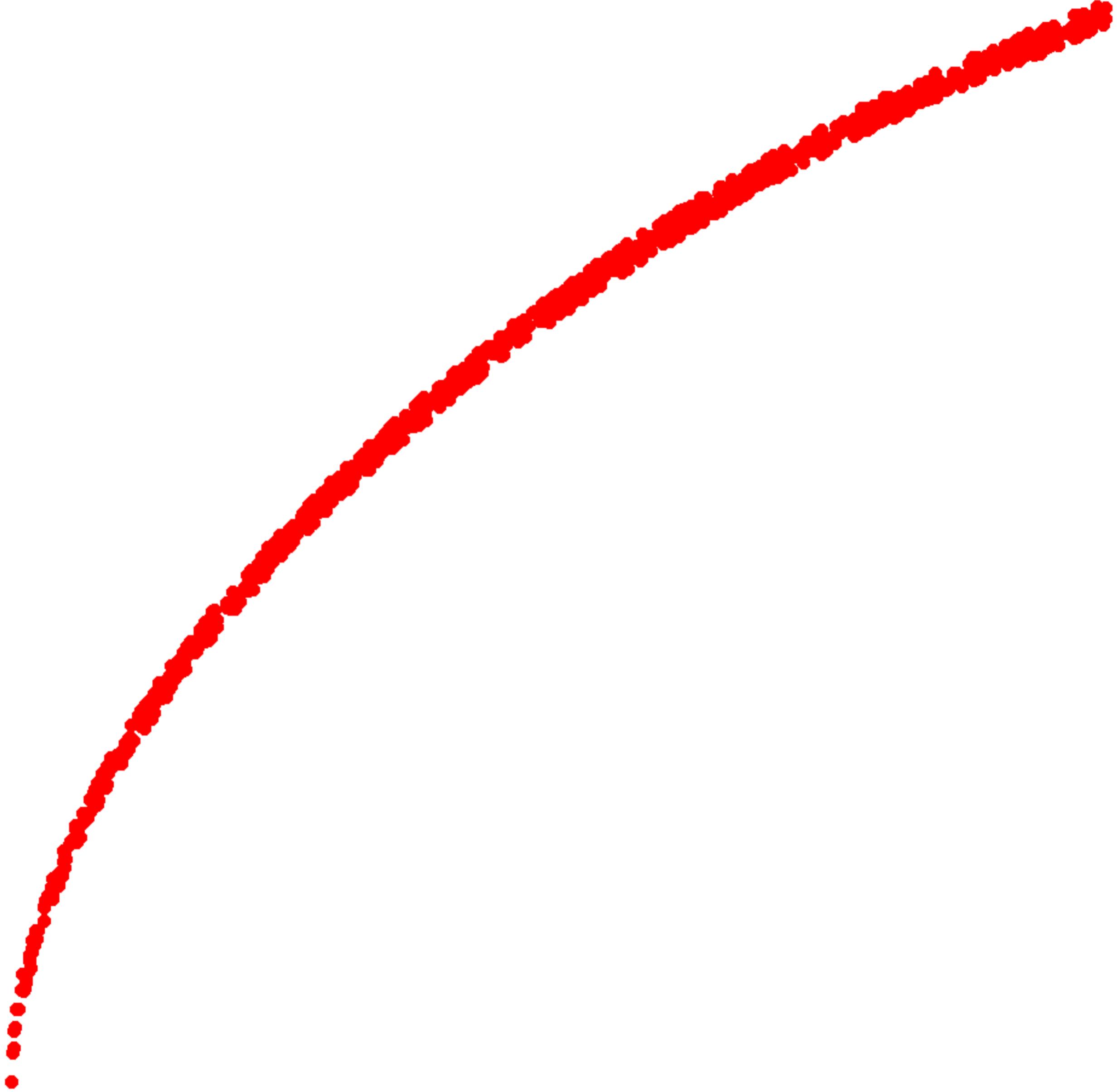
0.99926



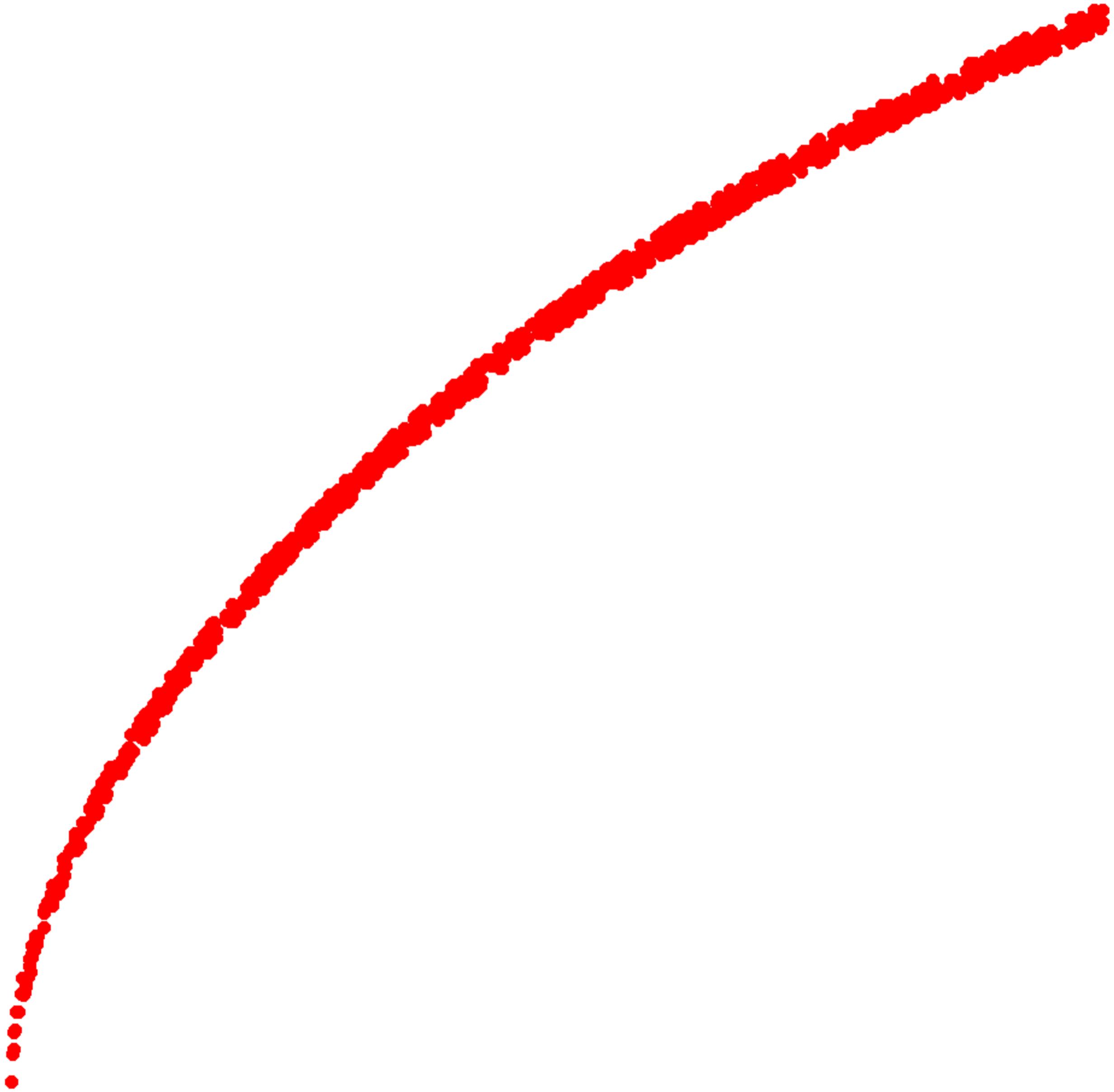
0.9993



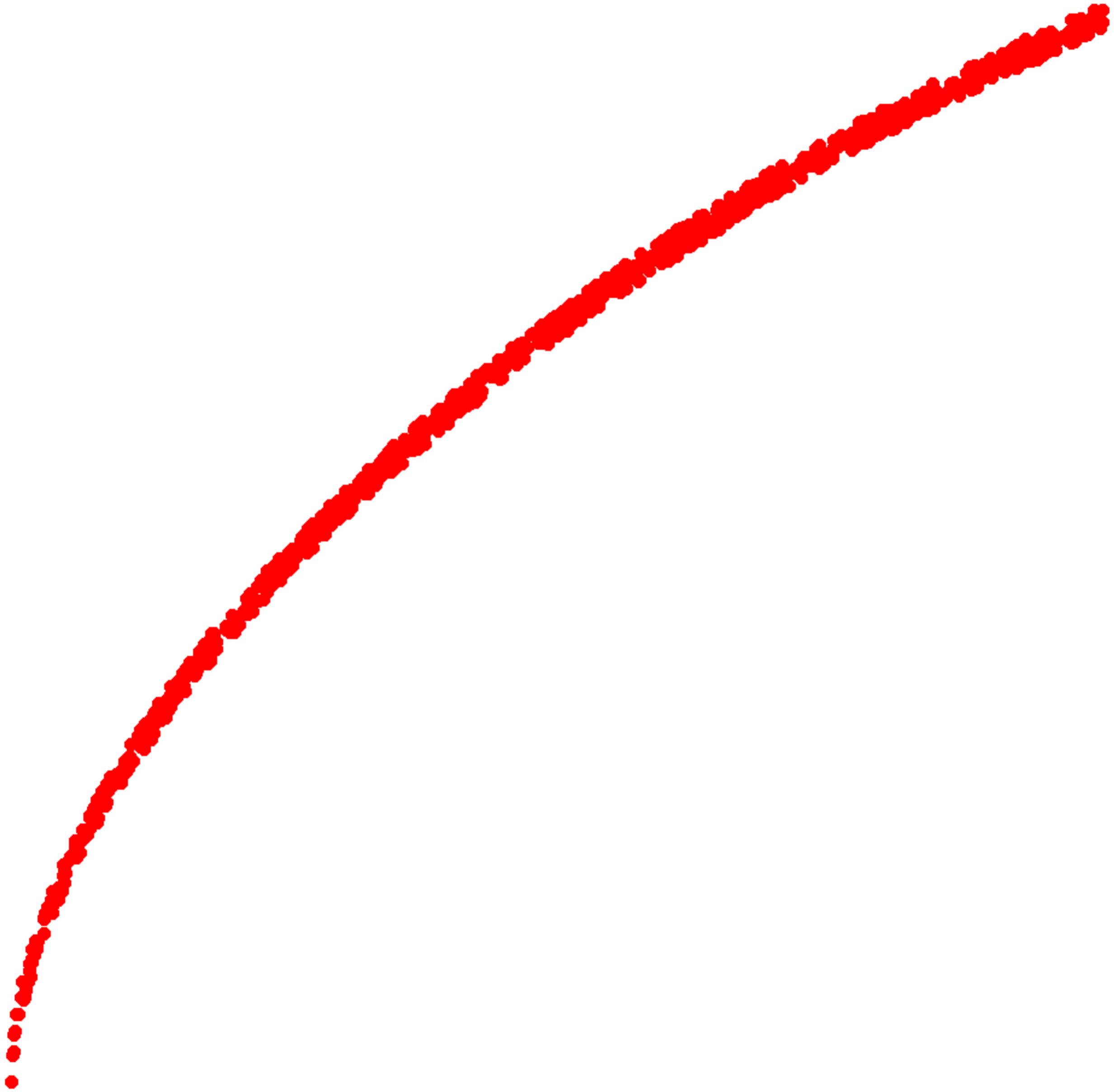
0.99933



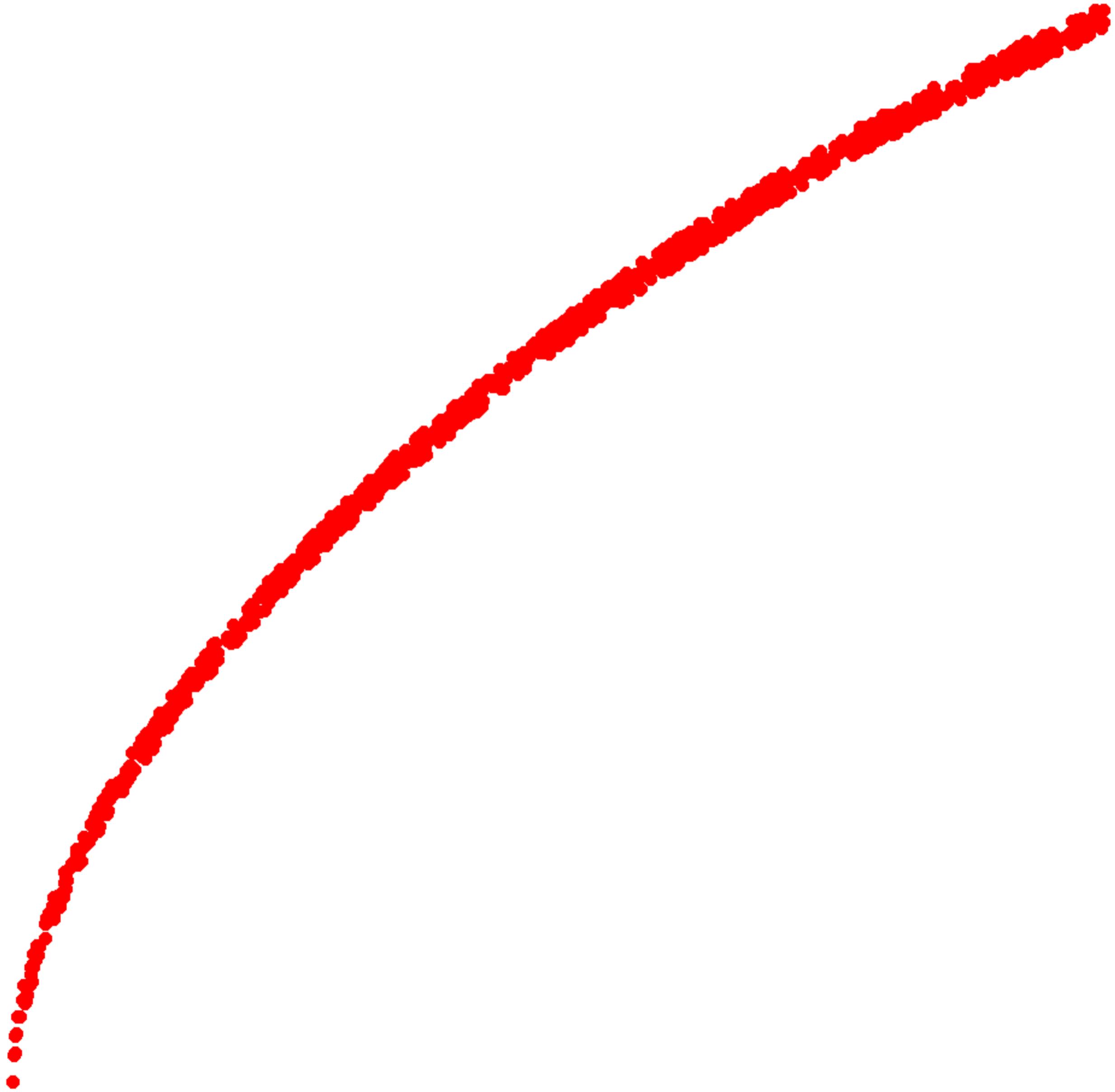
0.99936



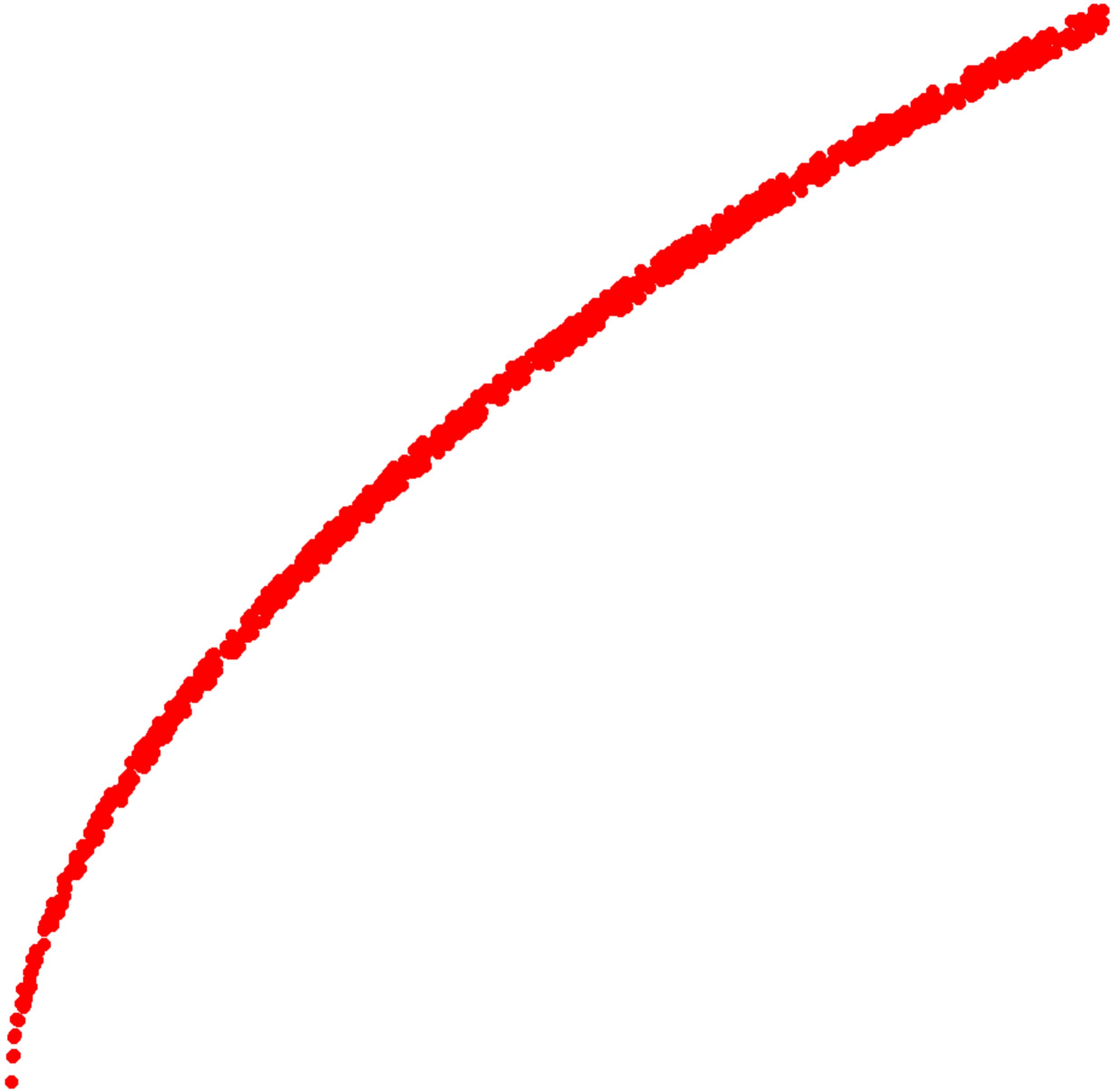
0.99938



0.9994



0.99942



0.99944



0.99946



0.99947



0.99948



0.9995



0.99951



0.99952



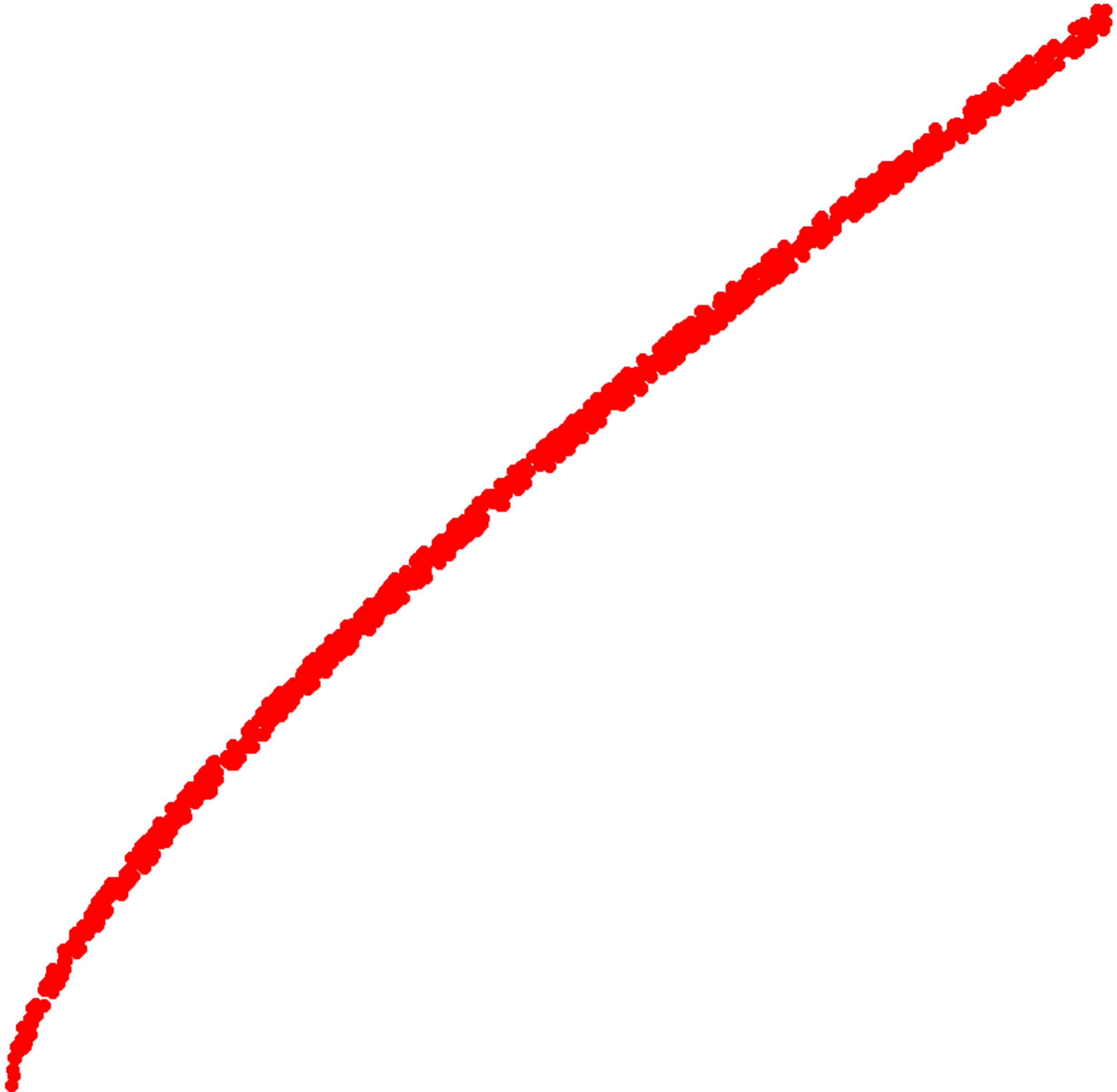
0.99952



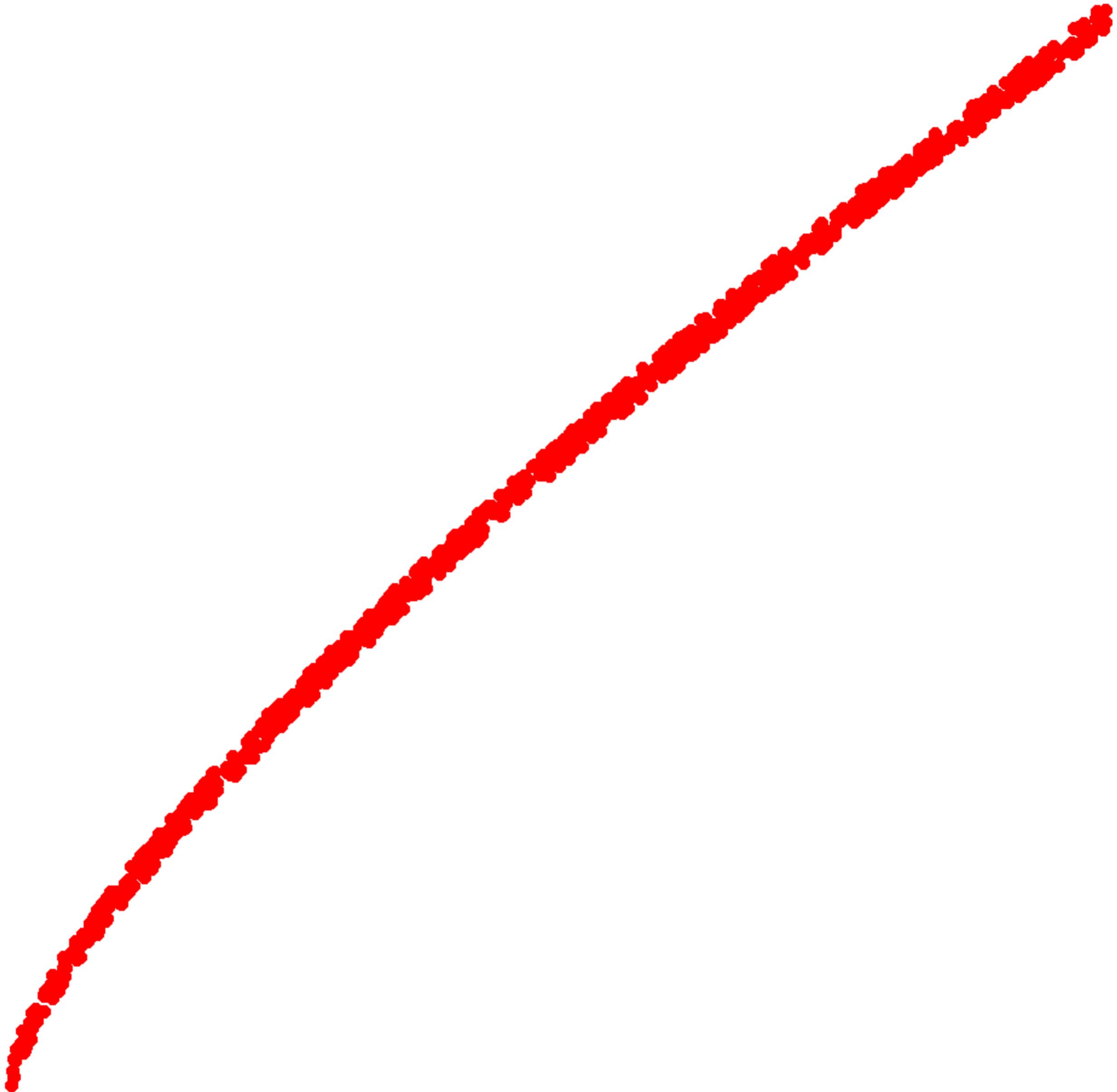
0.99953



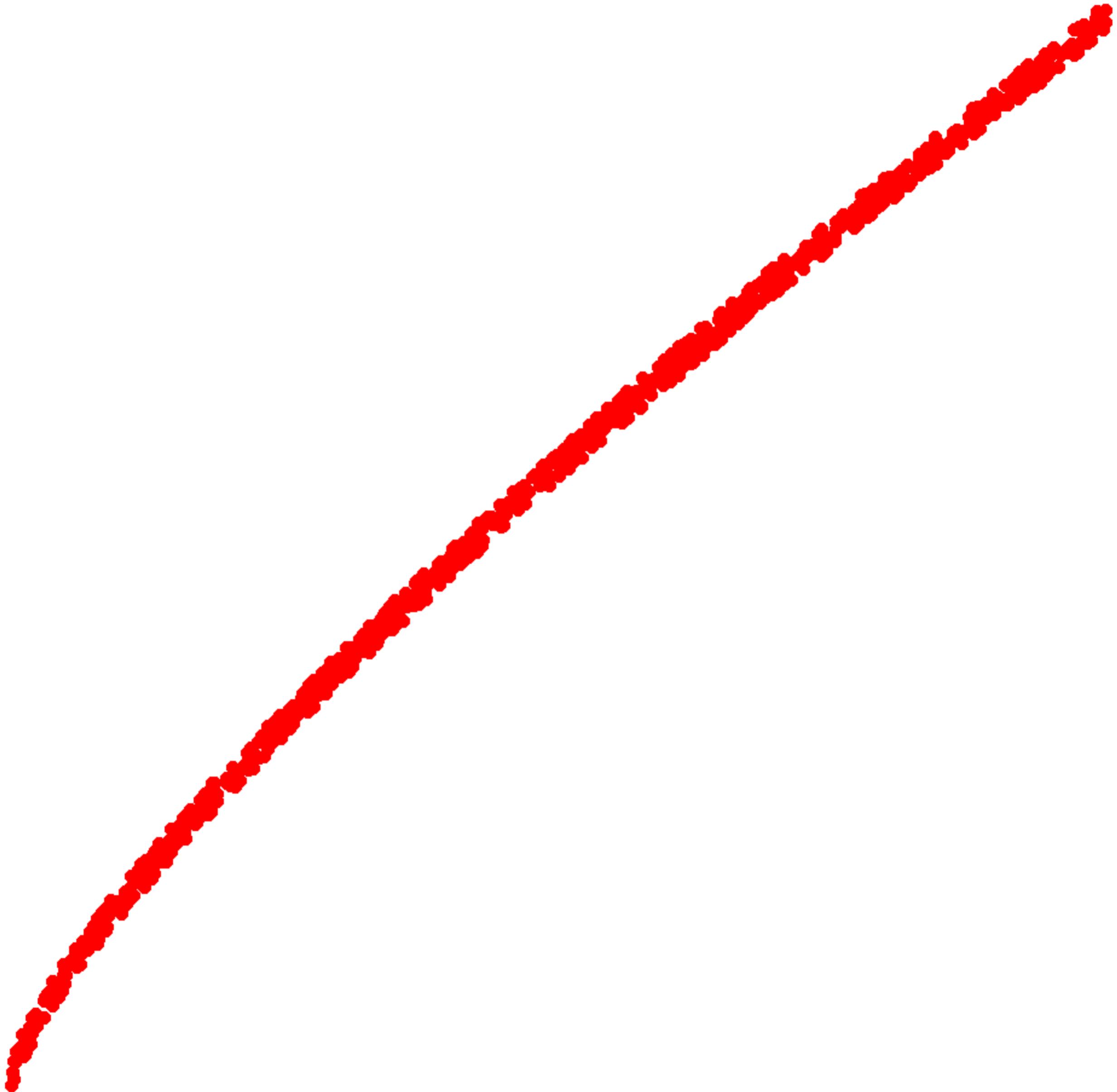
0.99954



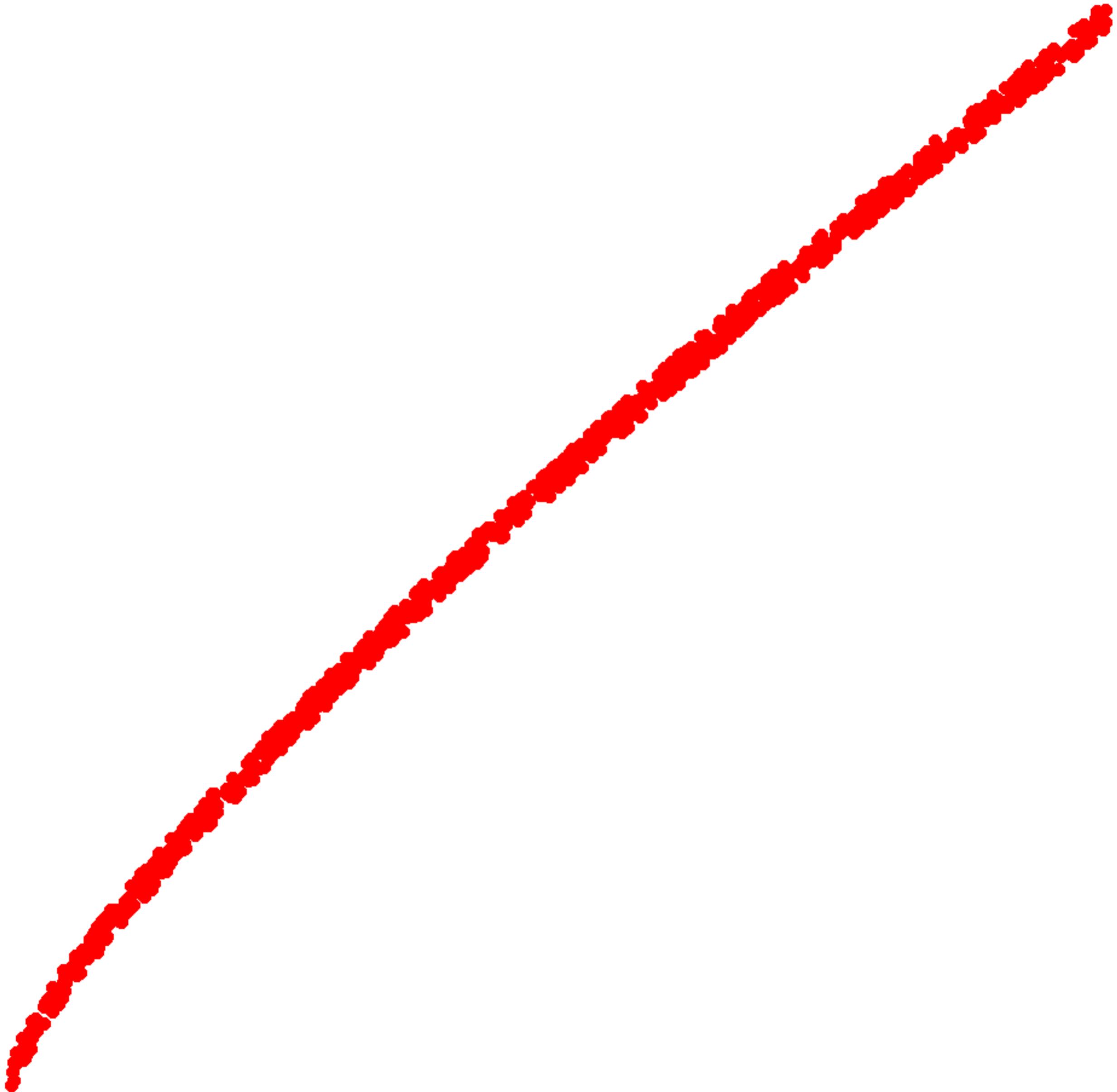
0.99954



0.99955



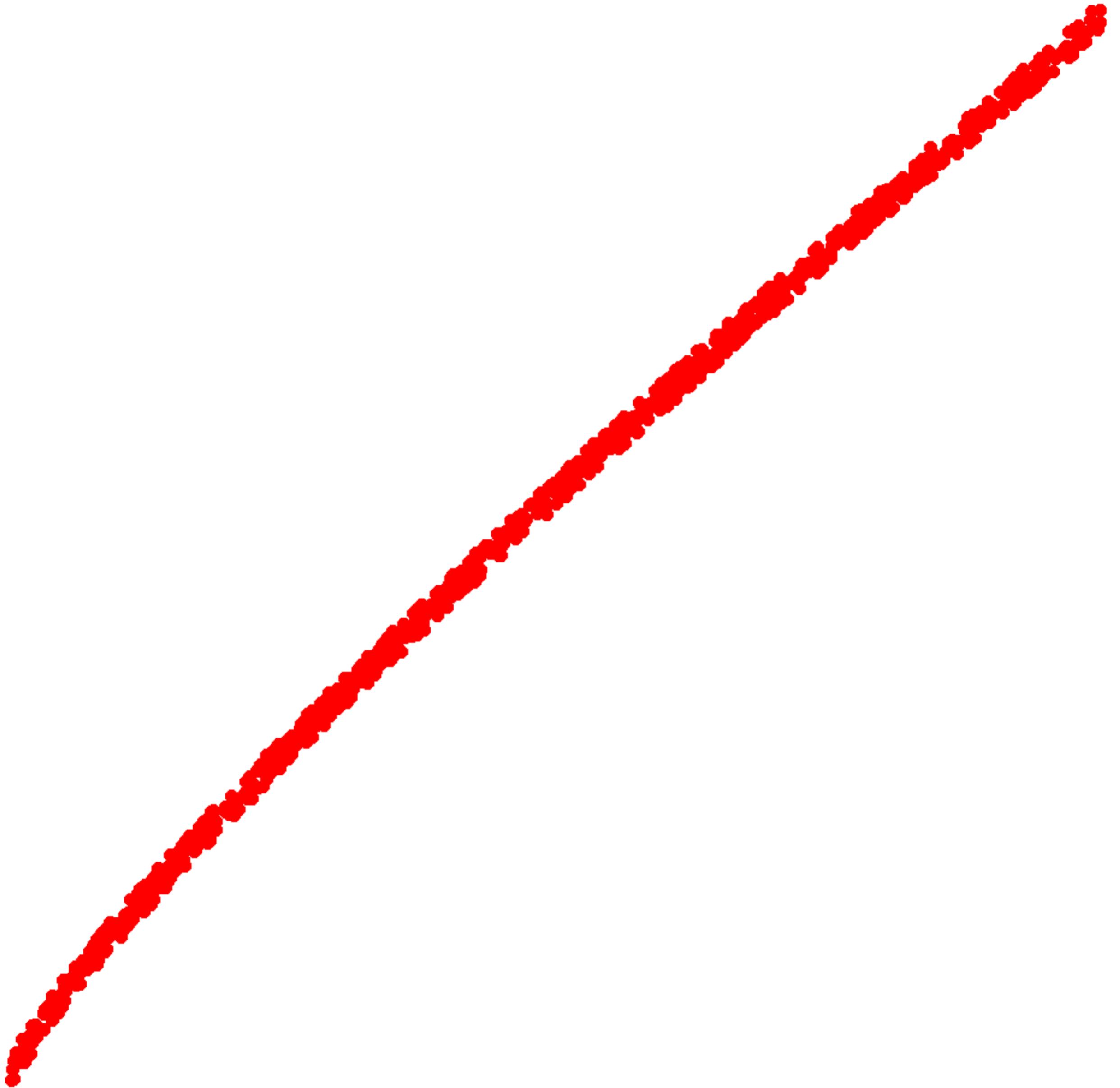
0.99955



0.99955



0.99956



0.99956



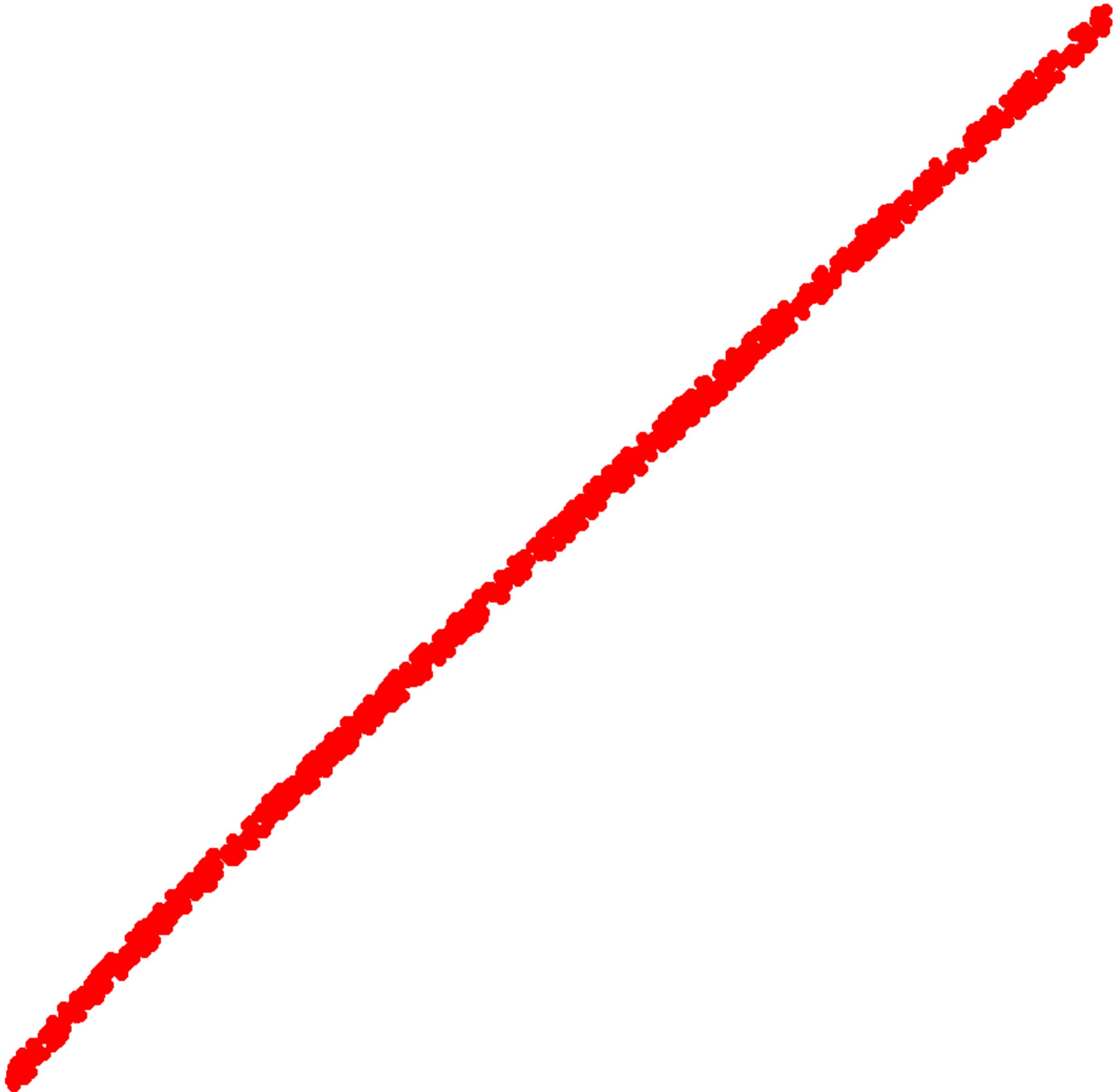
0.99956



0.99956



0.99956



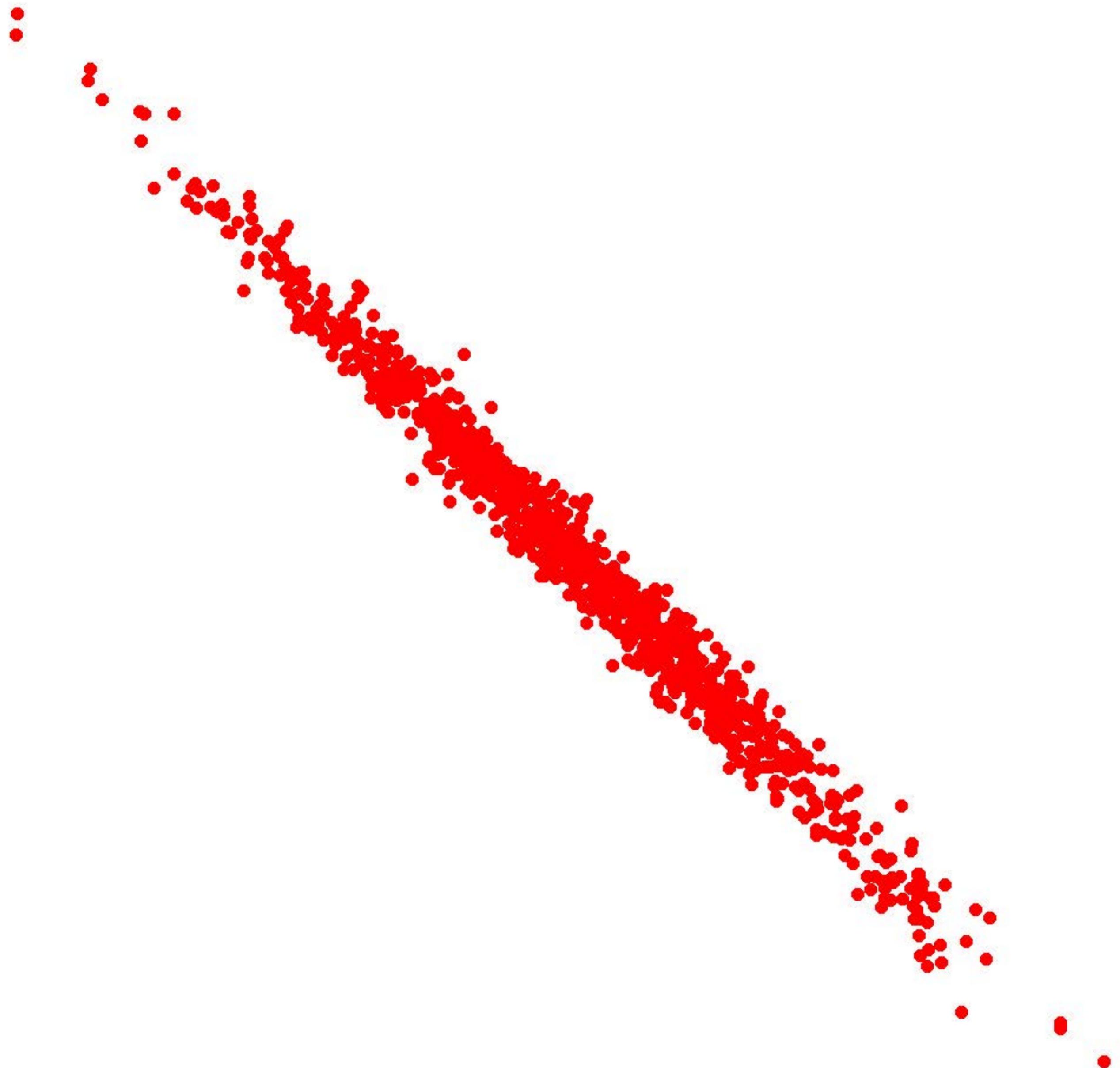
0.99956



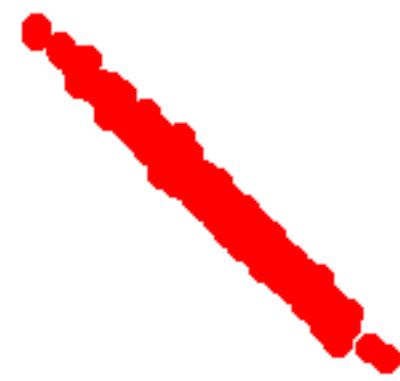
0.99955



-0.9892



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



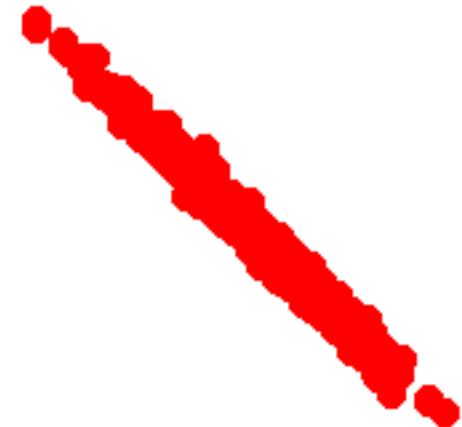
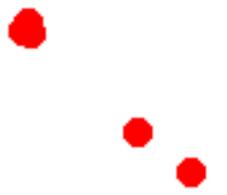
-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



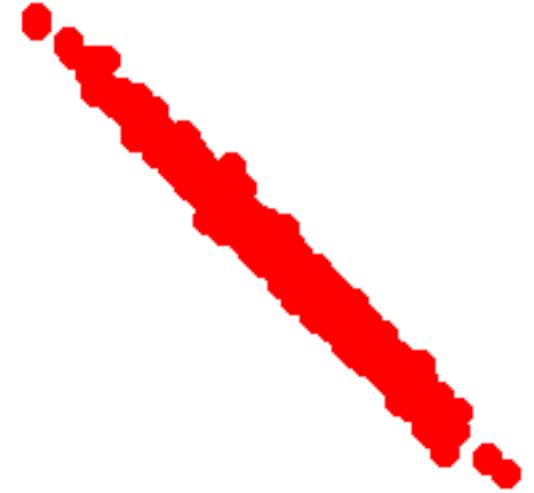
-0.95949



-0.95949



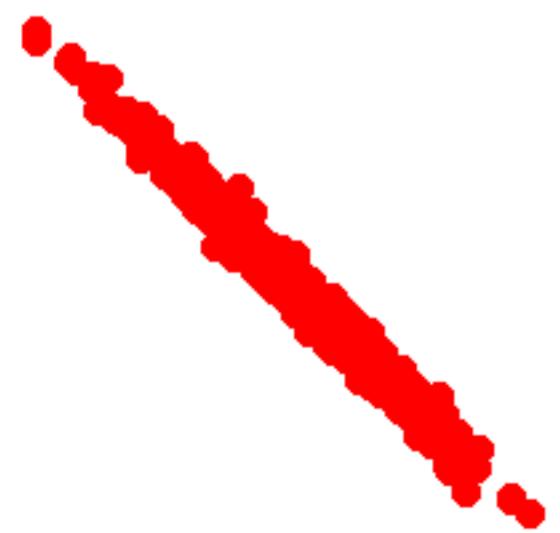
-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



-0.95949



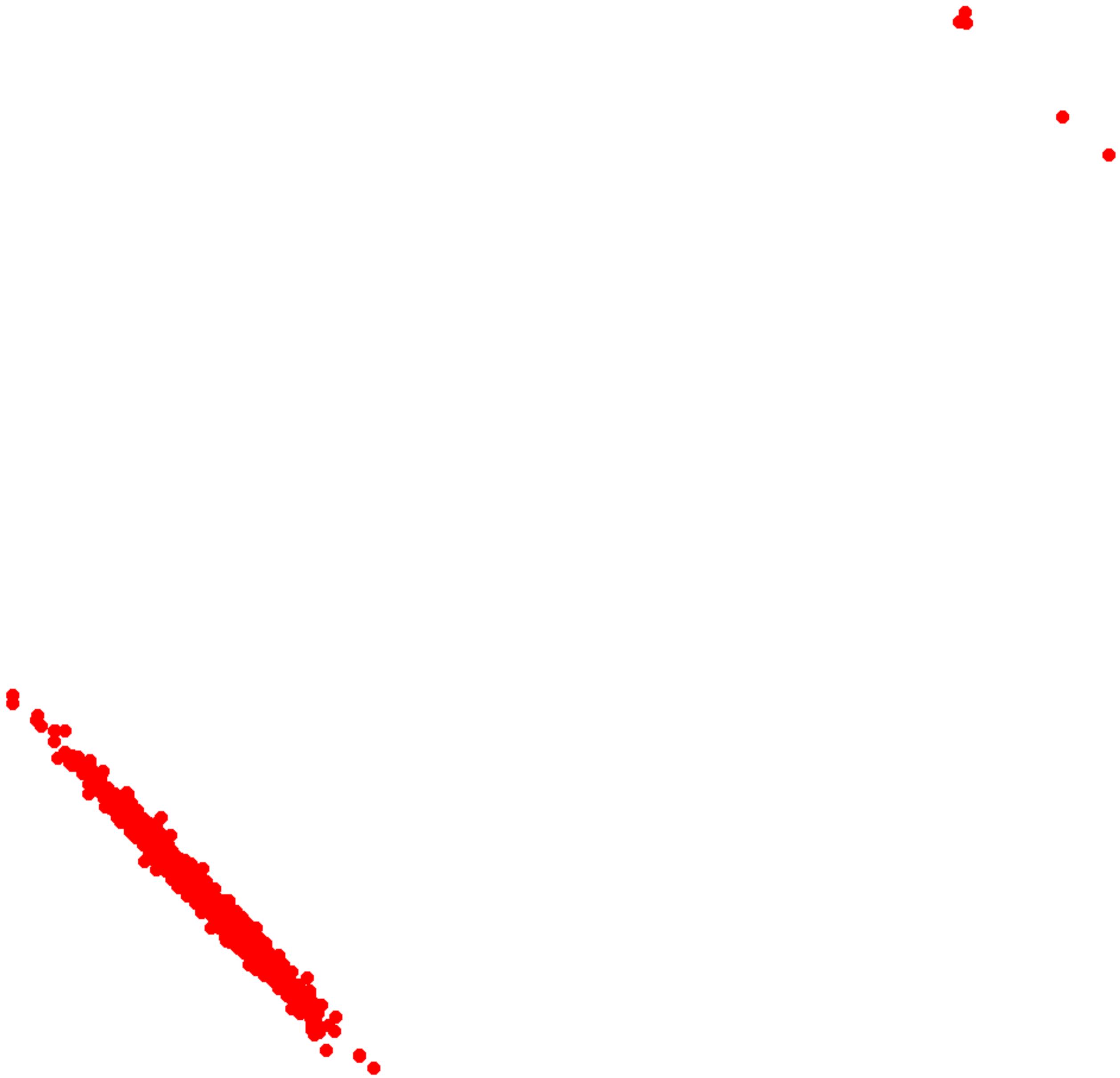
-0.95949



-0.95949



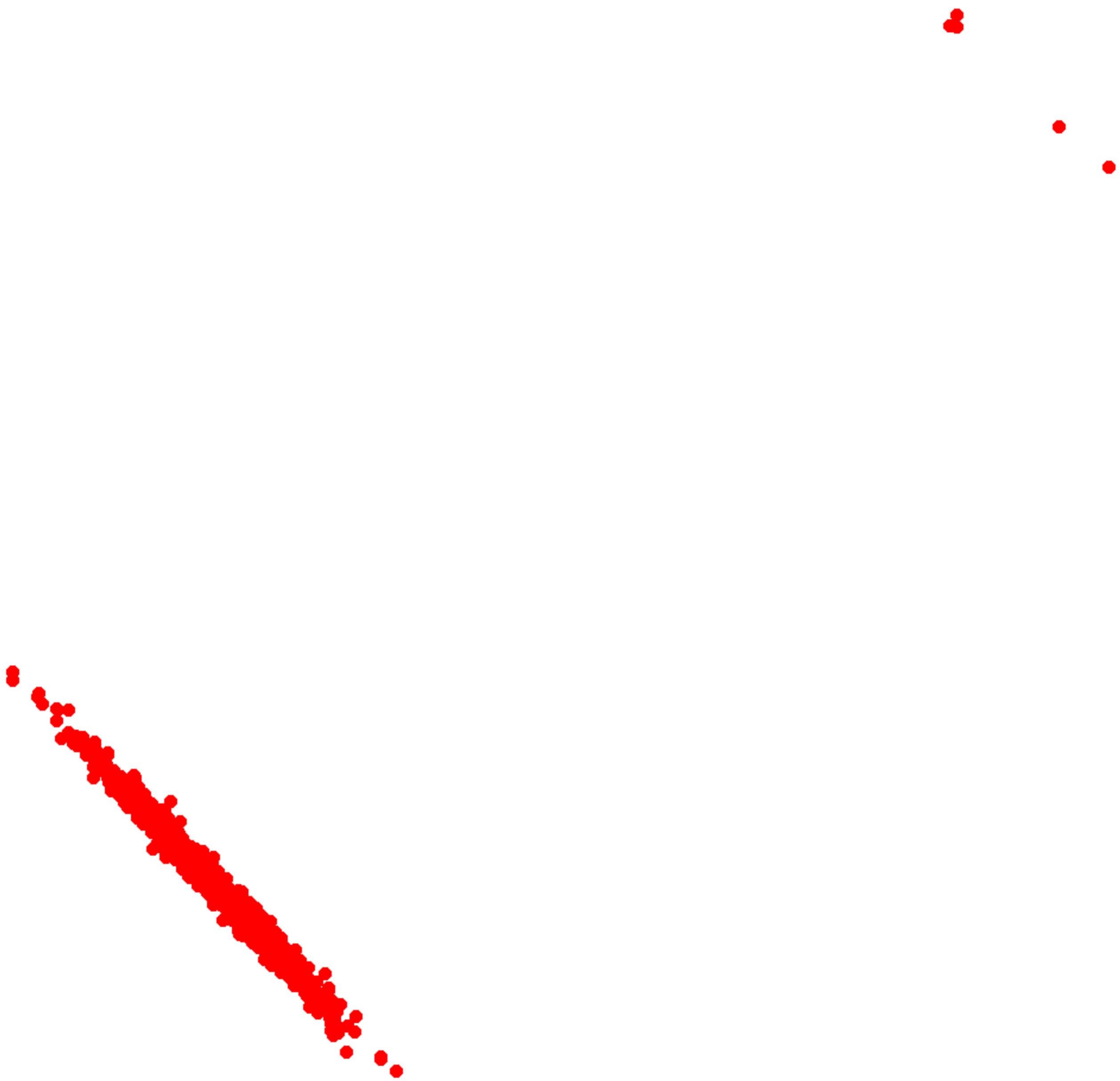
-0.95949



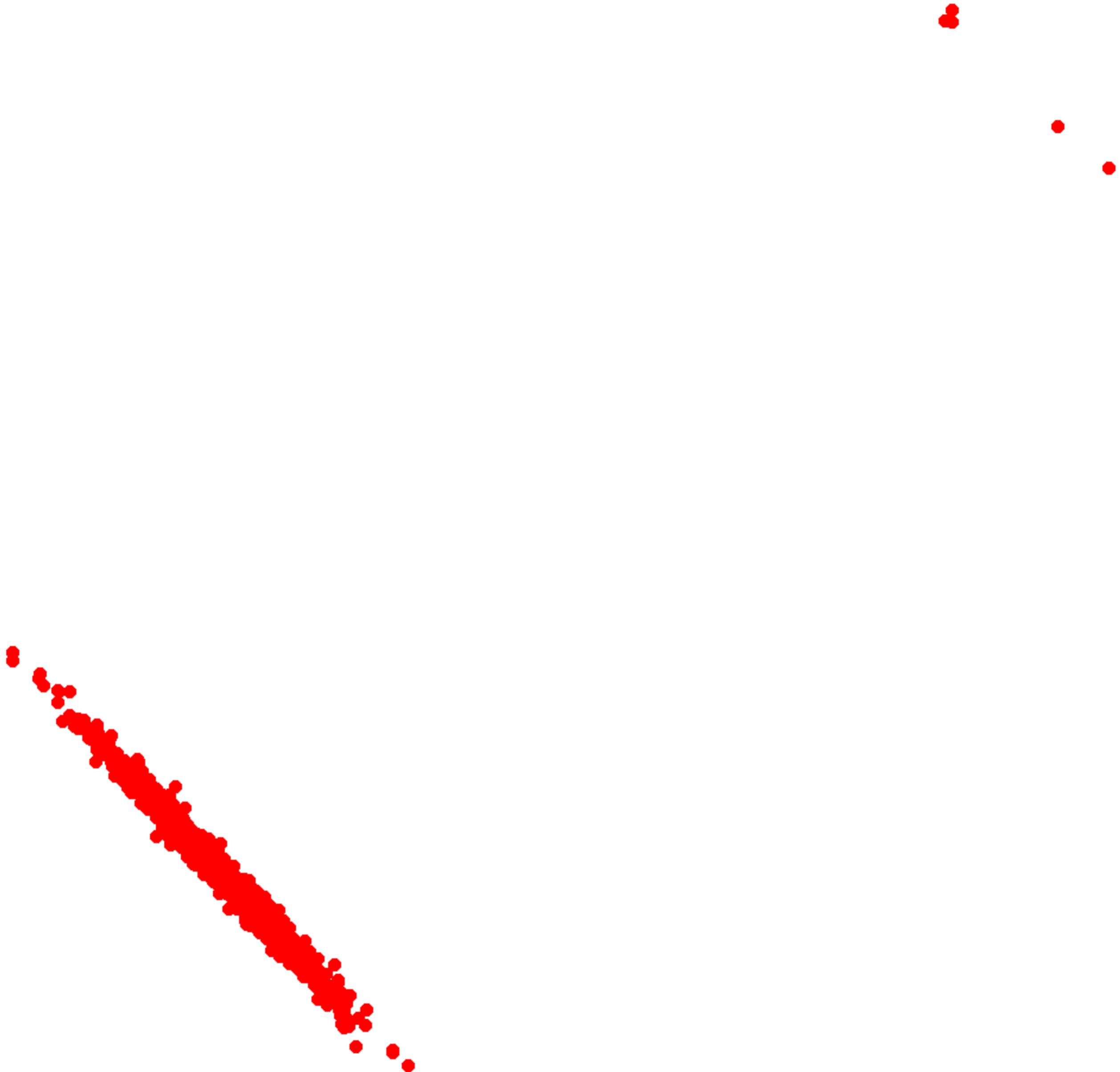
-0.95949



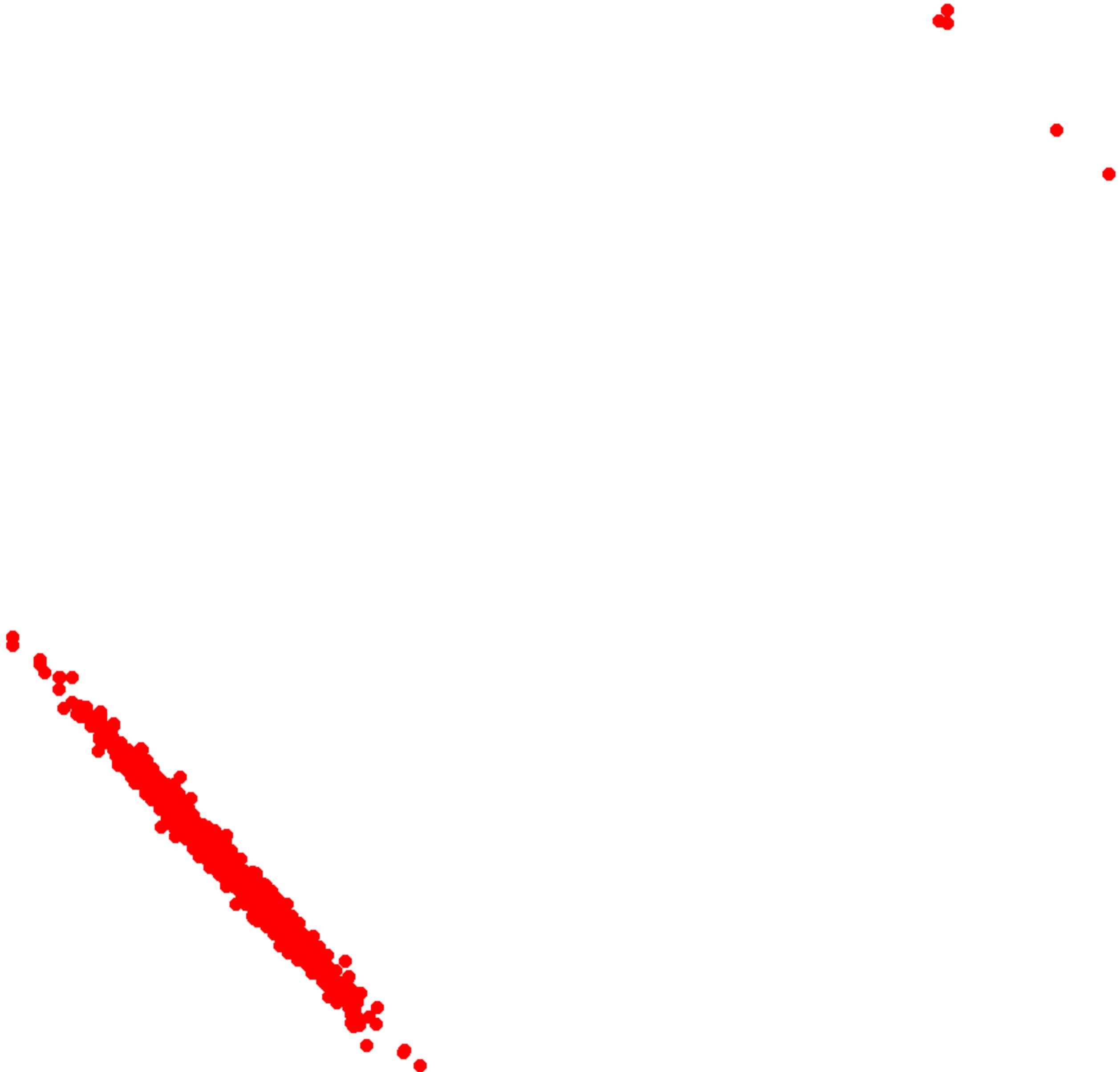
-0.95949



-0.95949



-0.95949



-0.95949



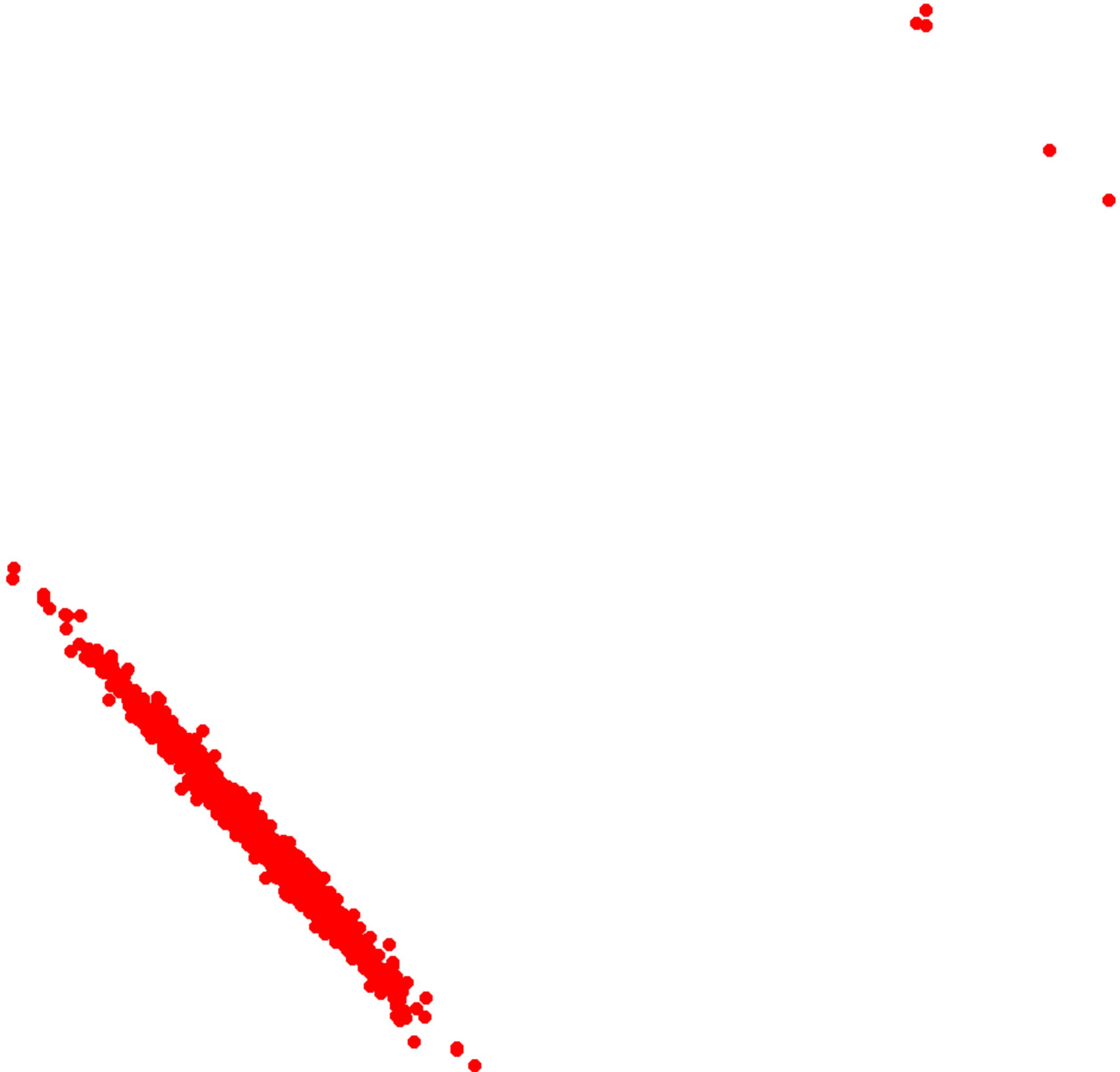
-0.95949



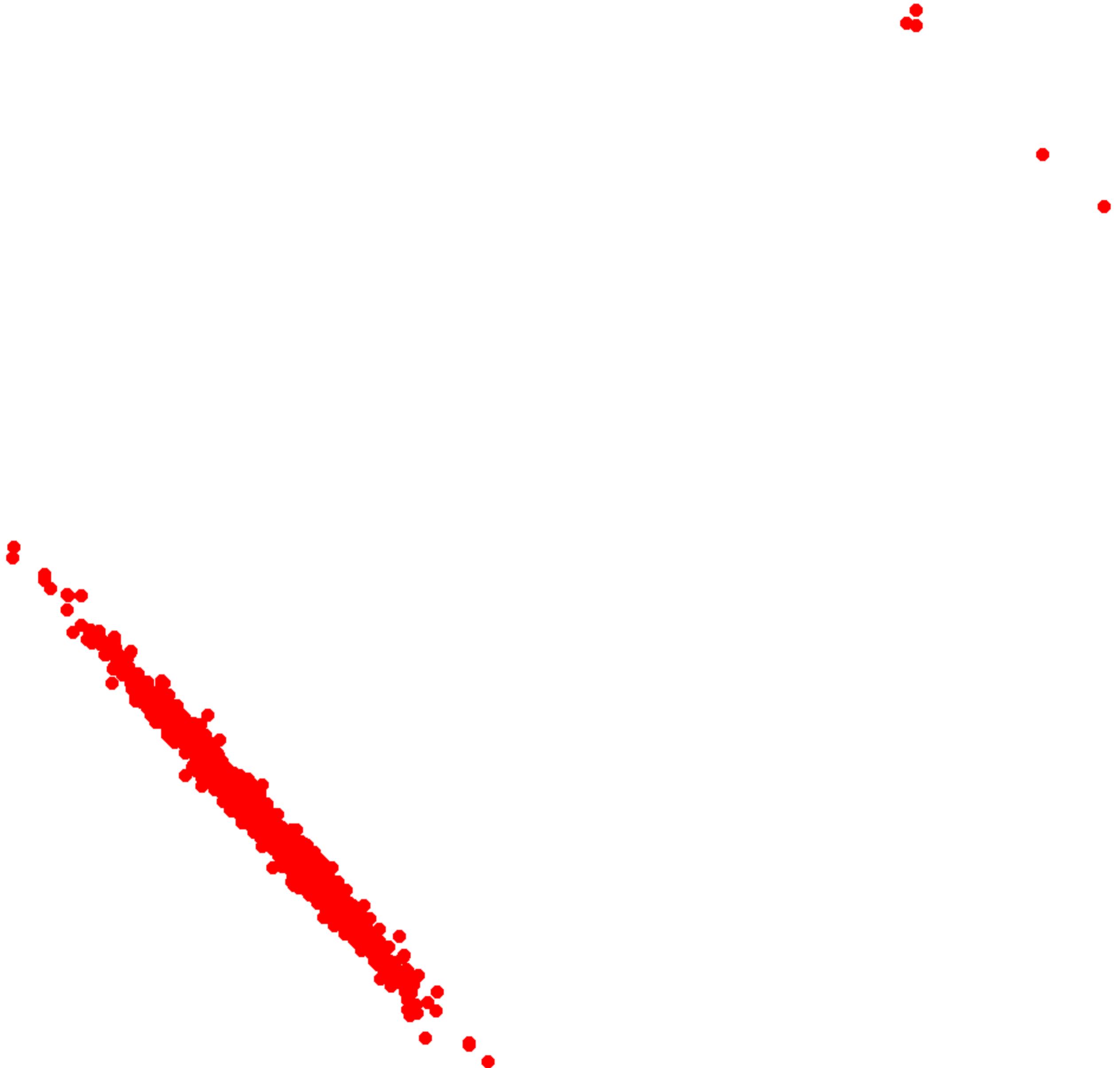
-0.95949



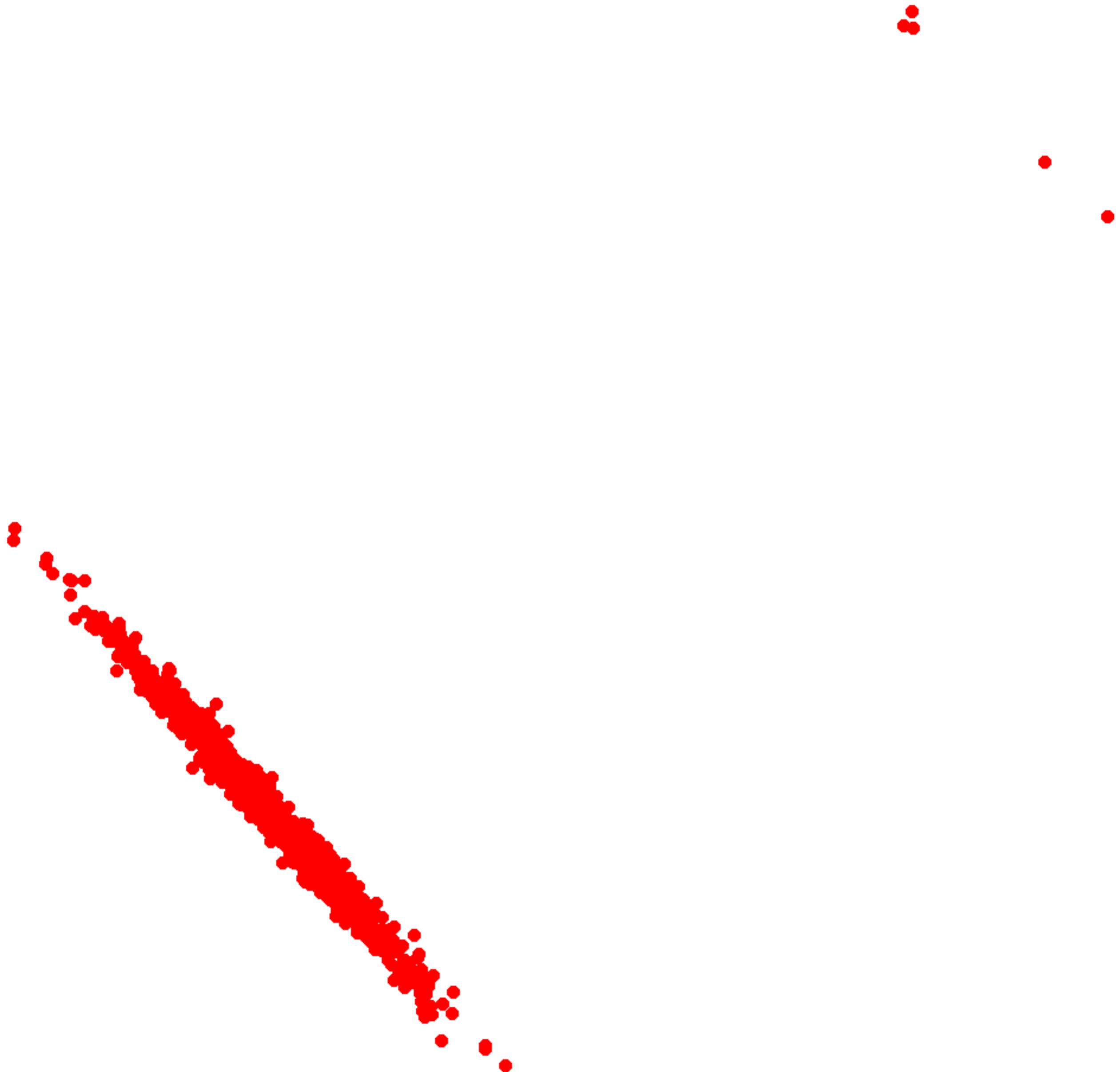
-0.95949



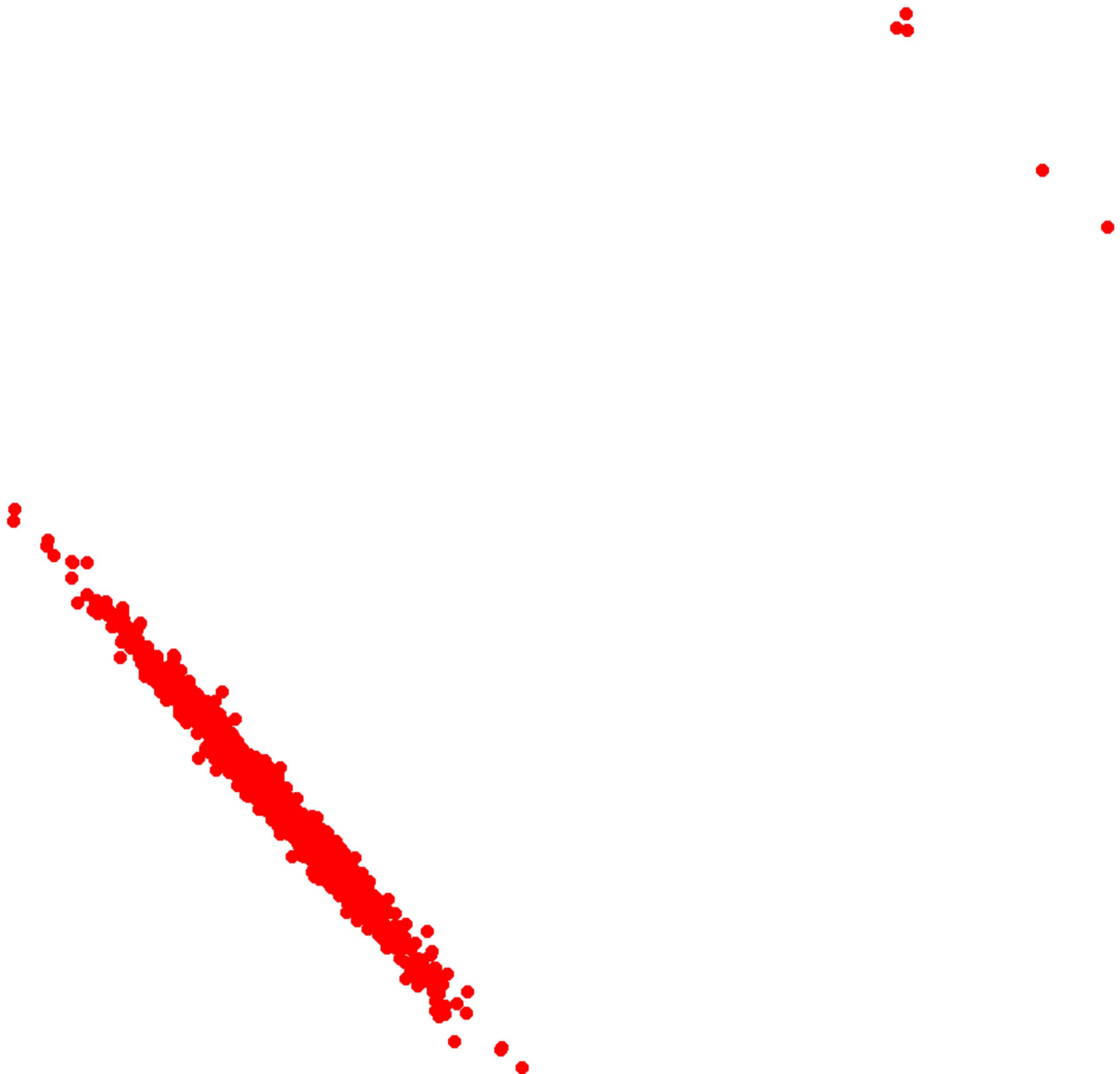
-0.95949



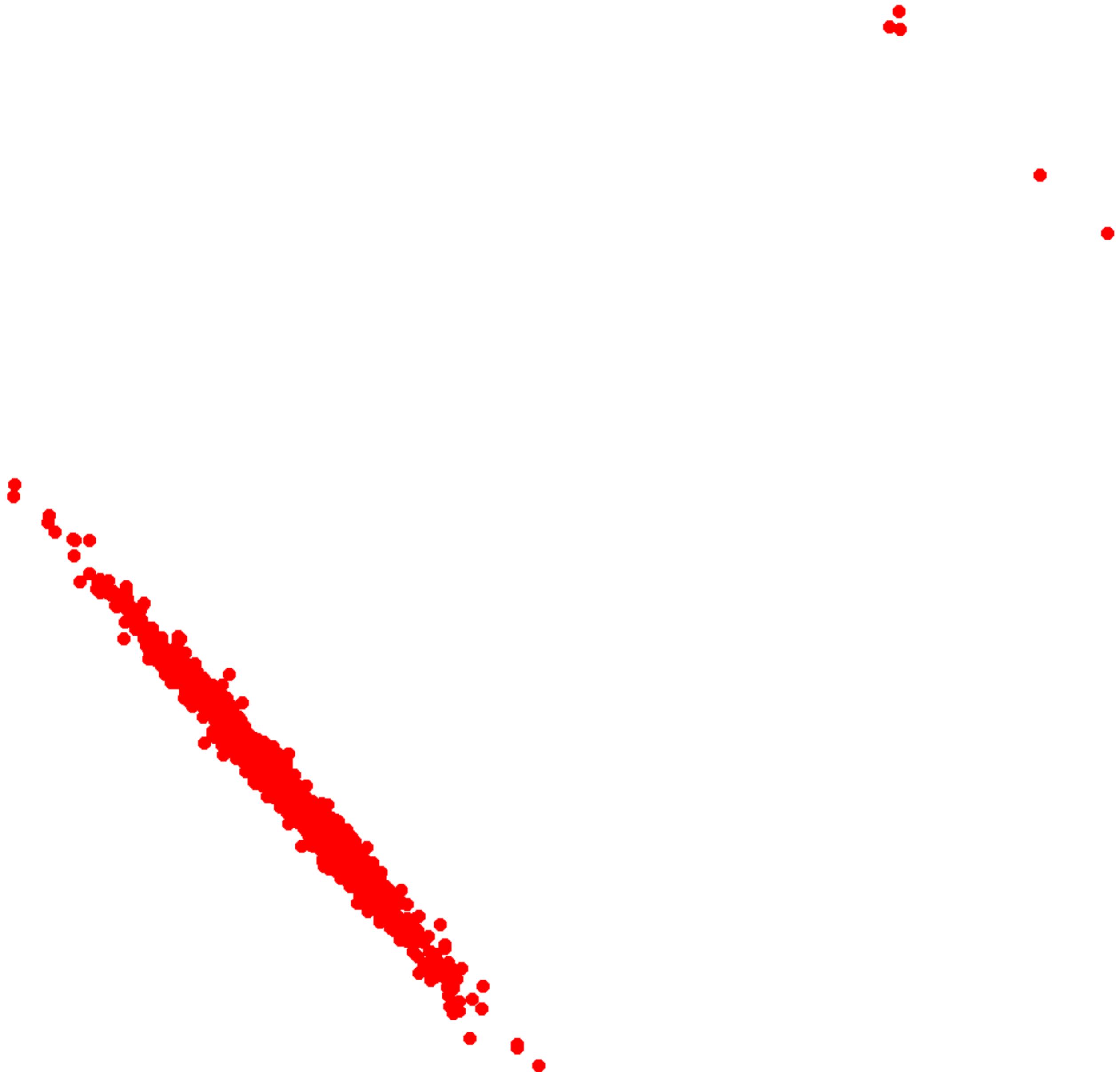
-0.95949



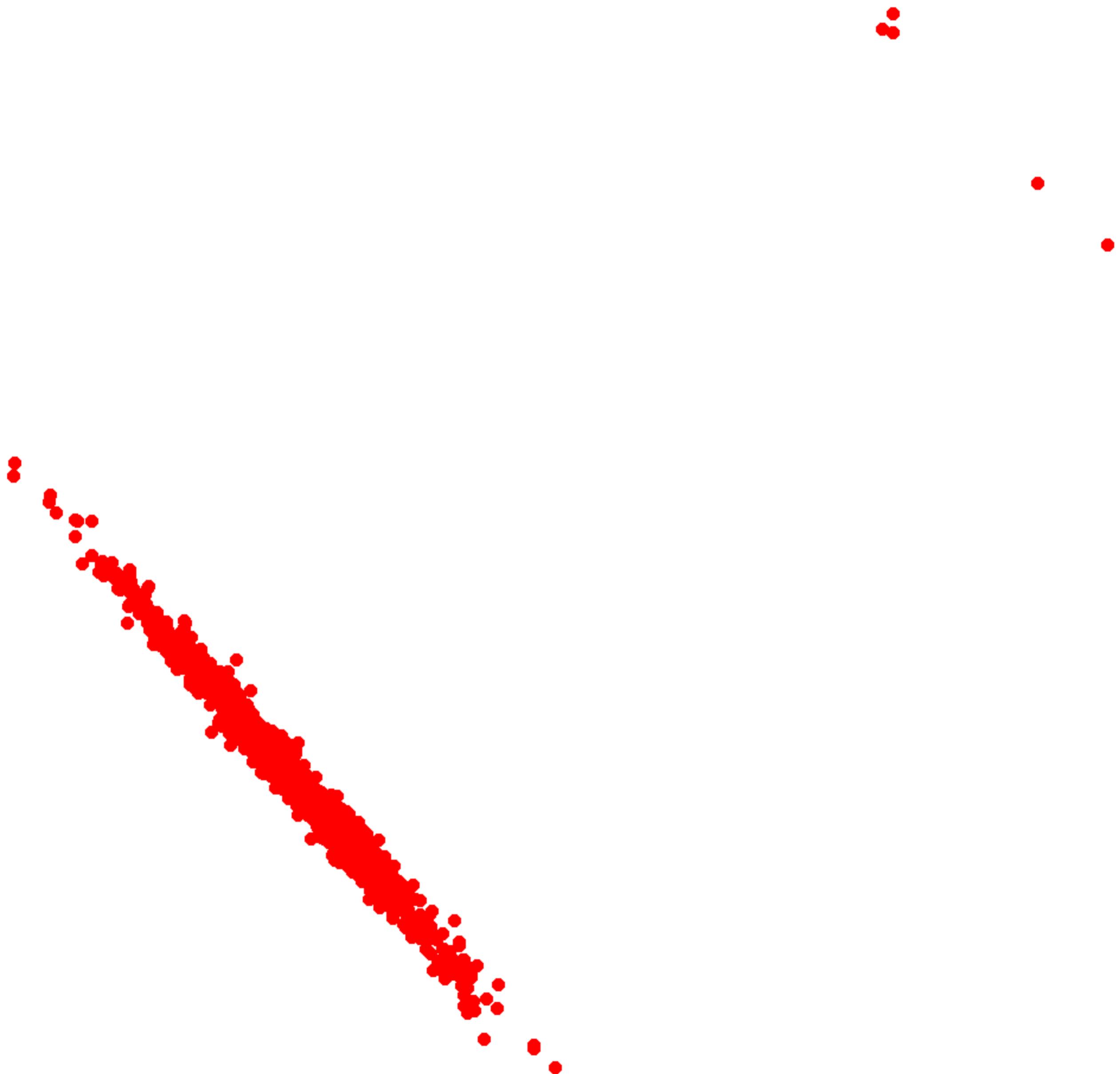
-0.95949



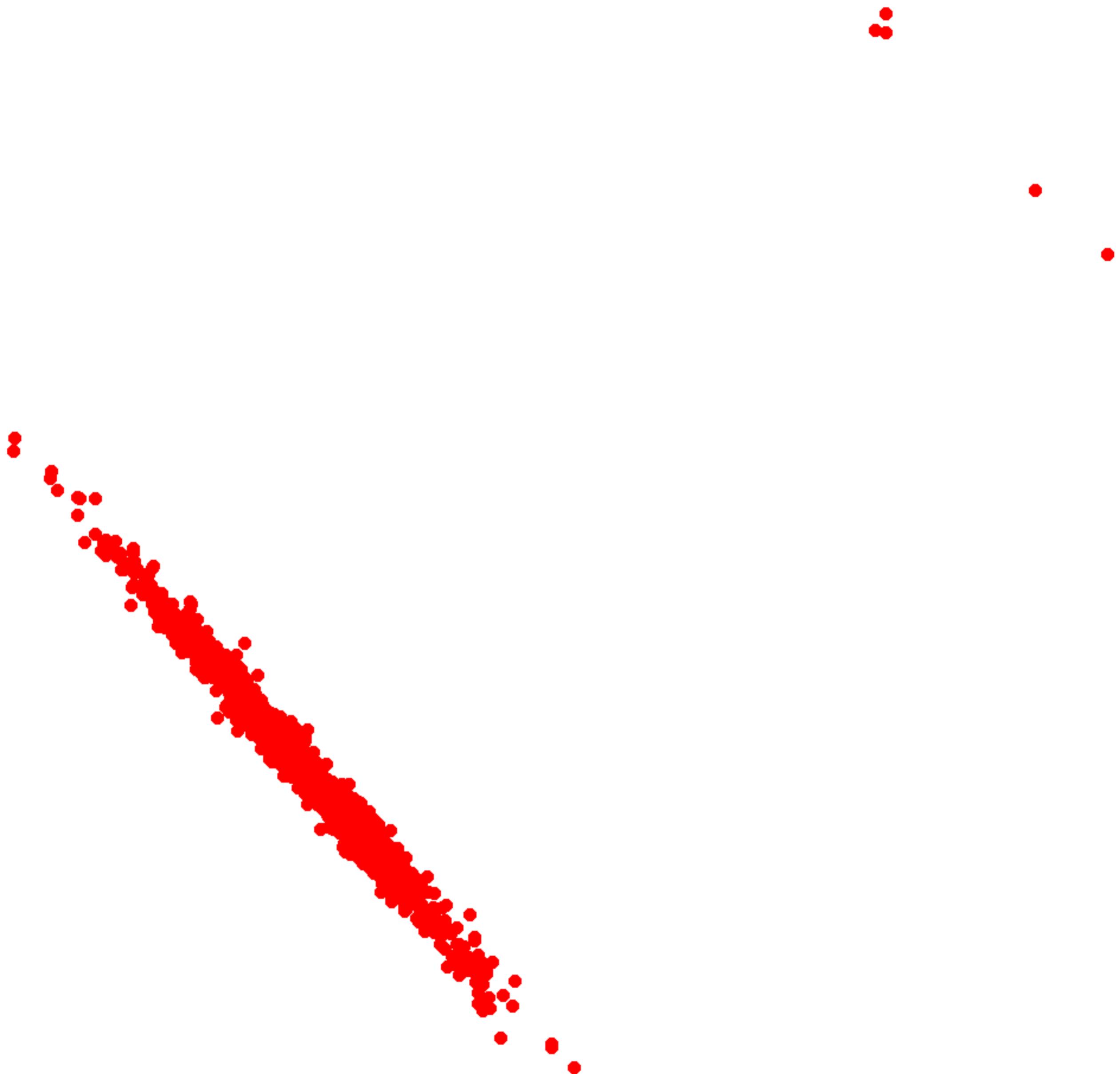
-0.95949



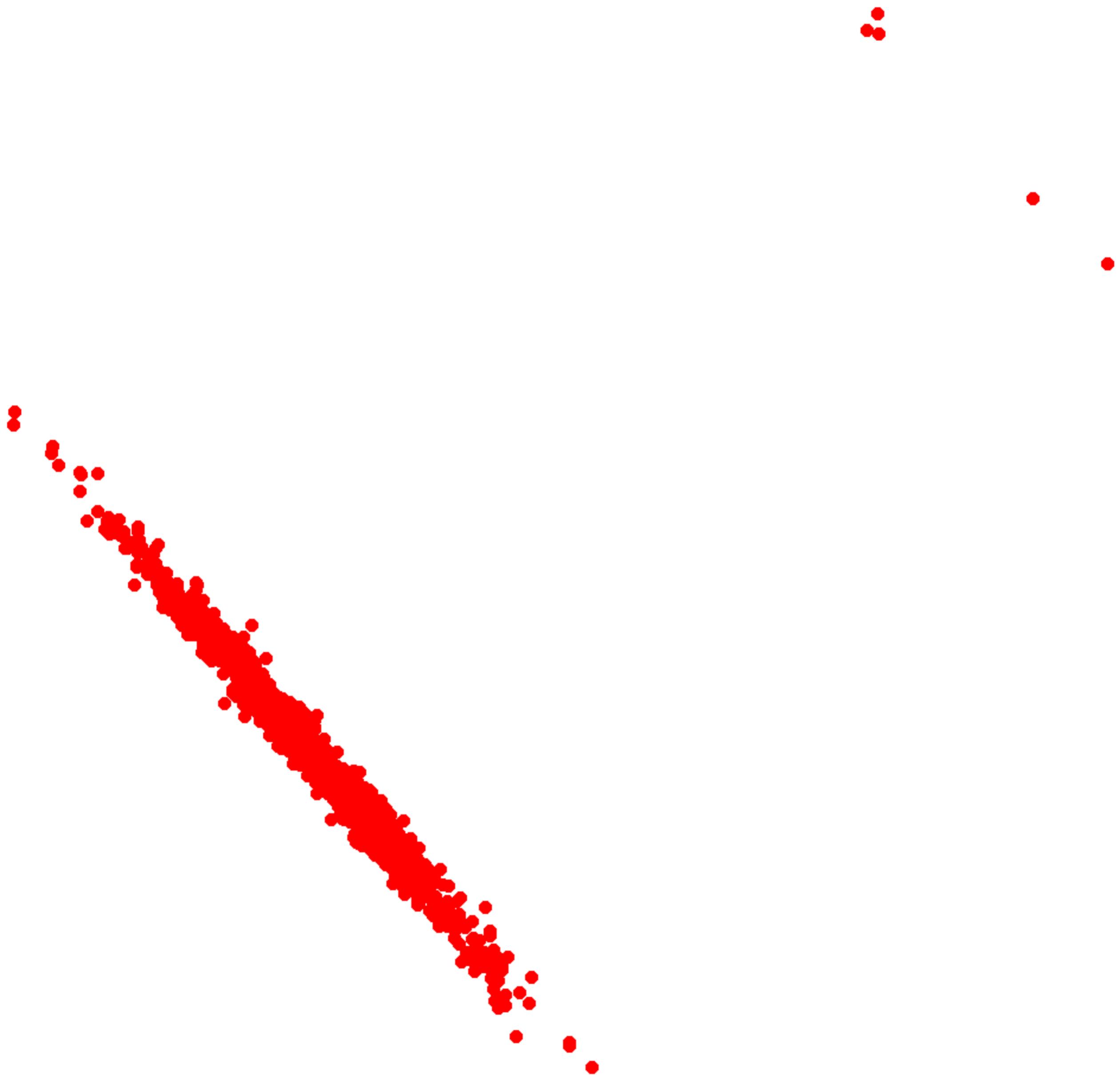
-0.95949



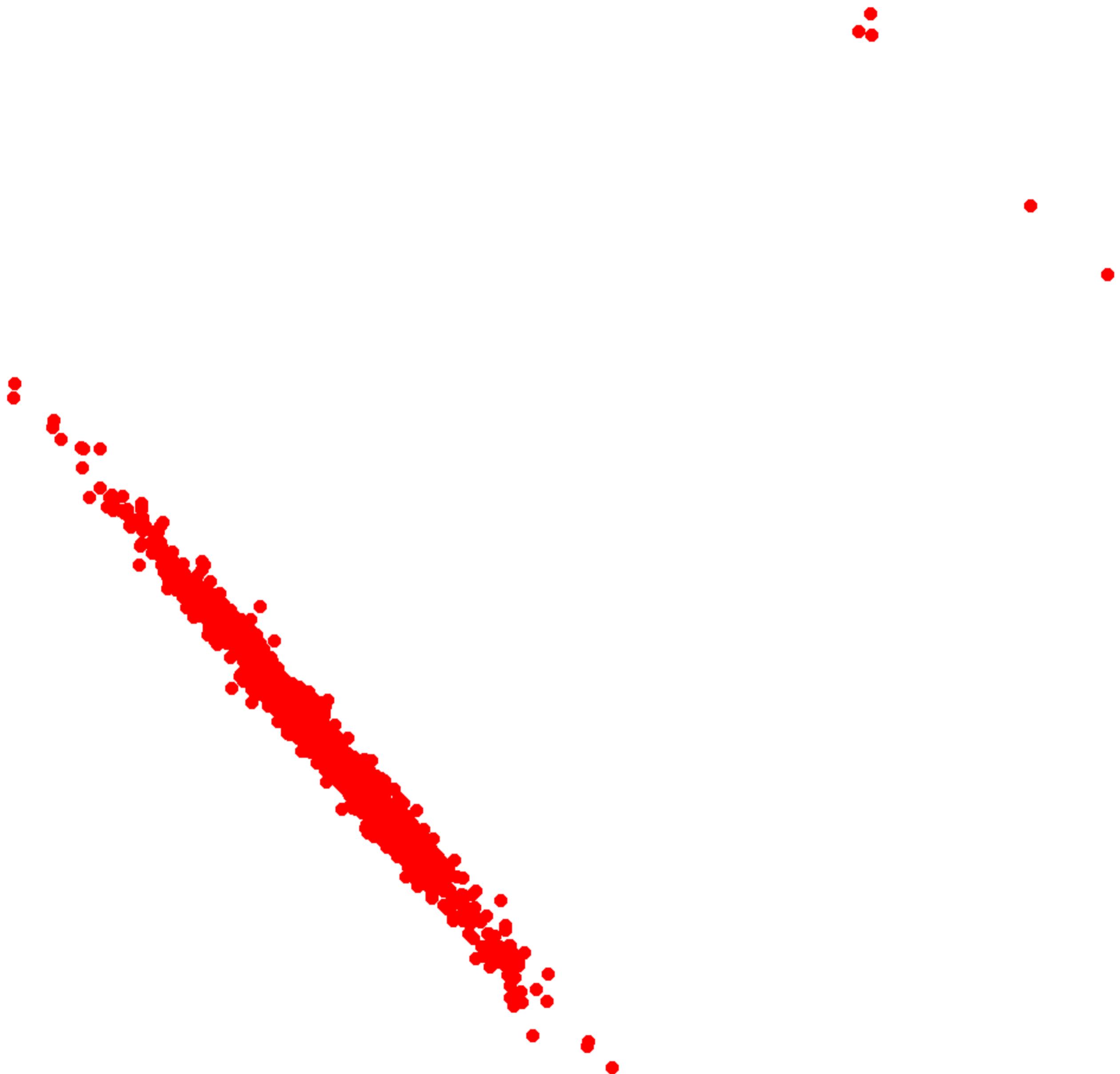
-0.95949



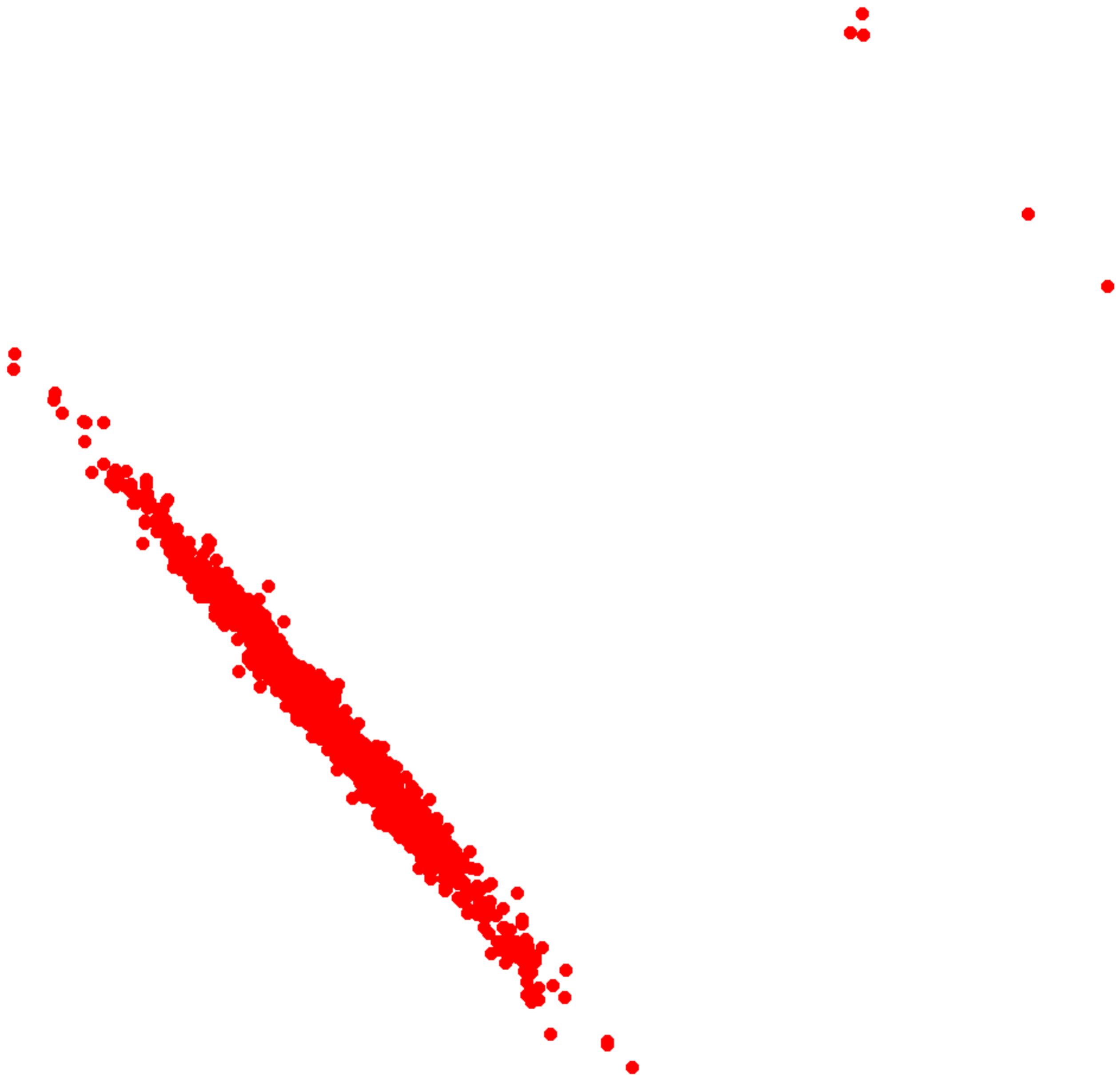
-0.95949



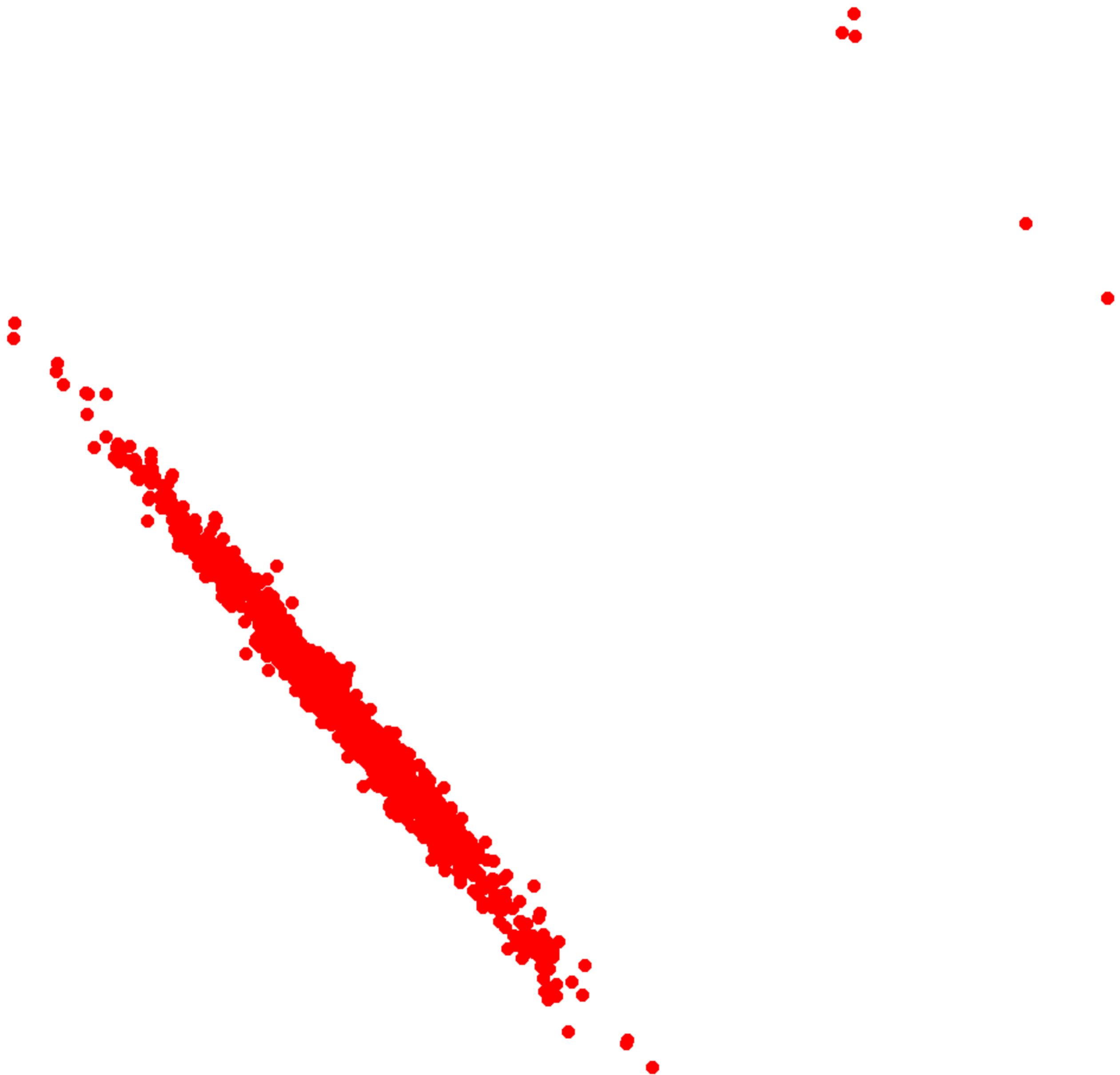
-0.95949



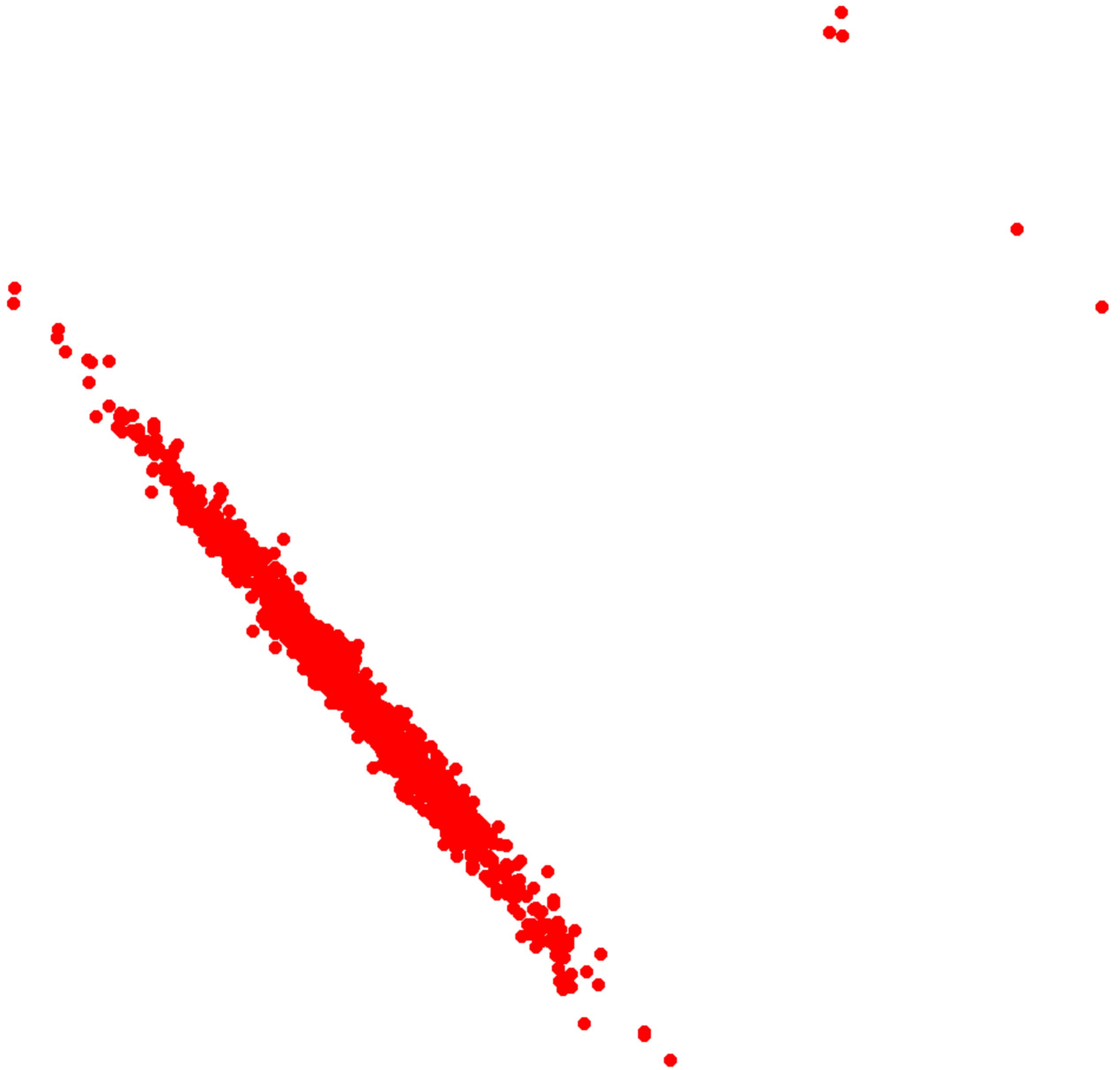
-0.95949



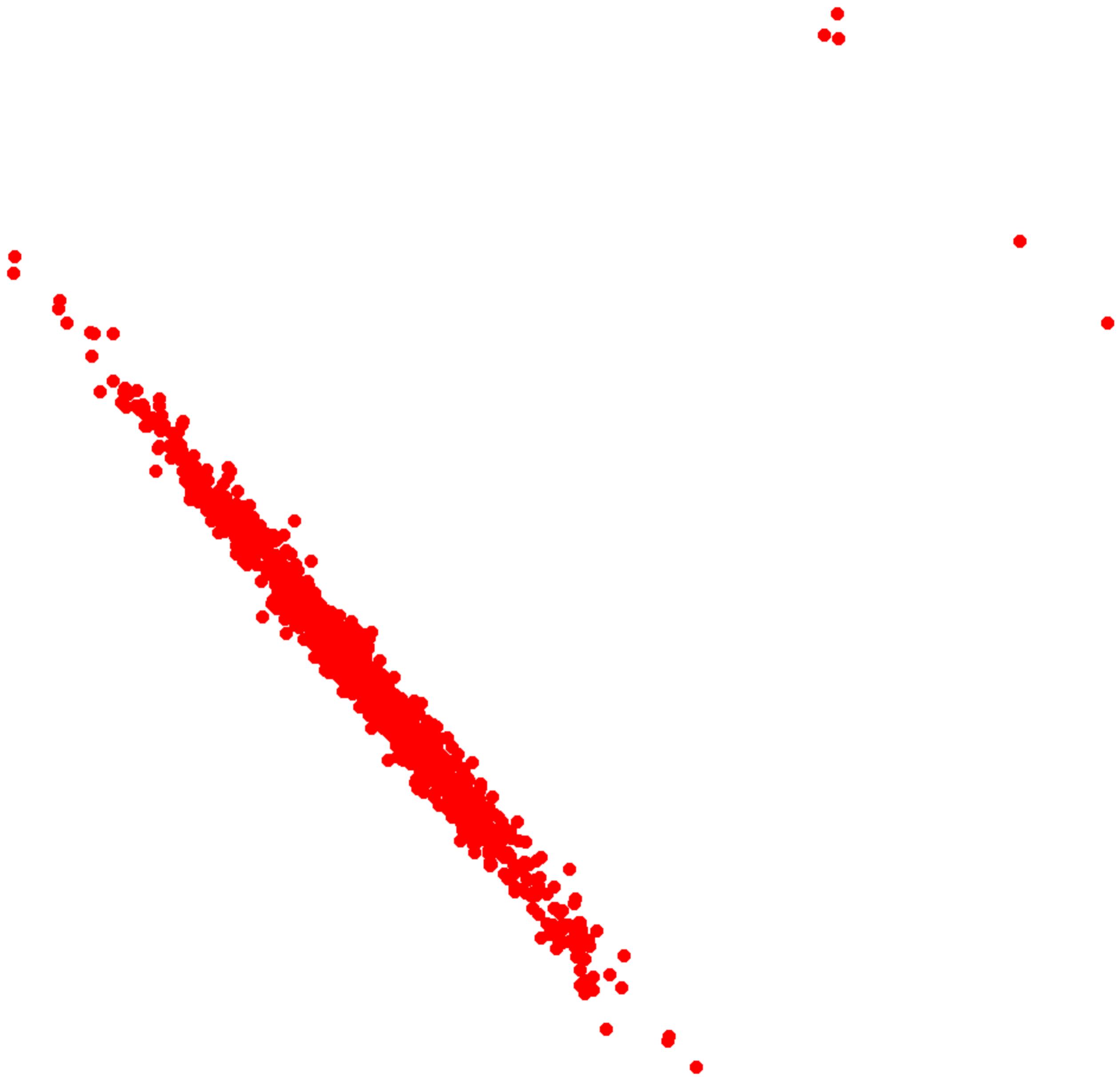
-0.95949



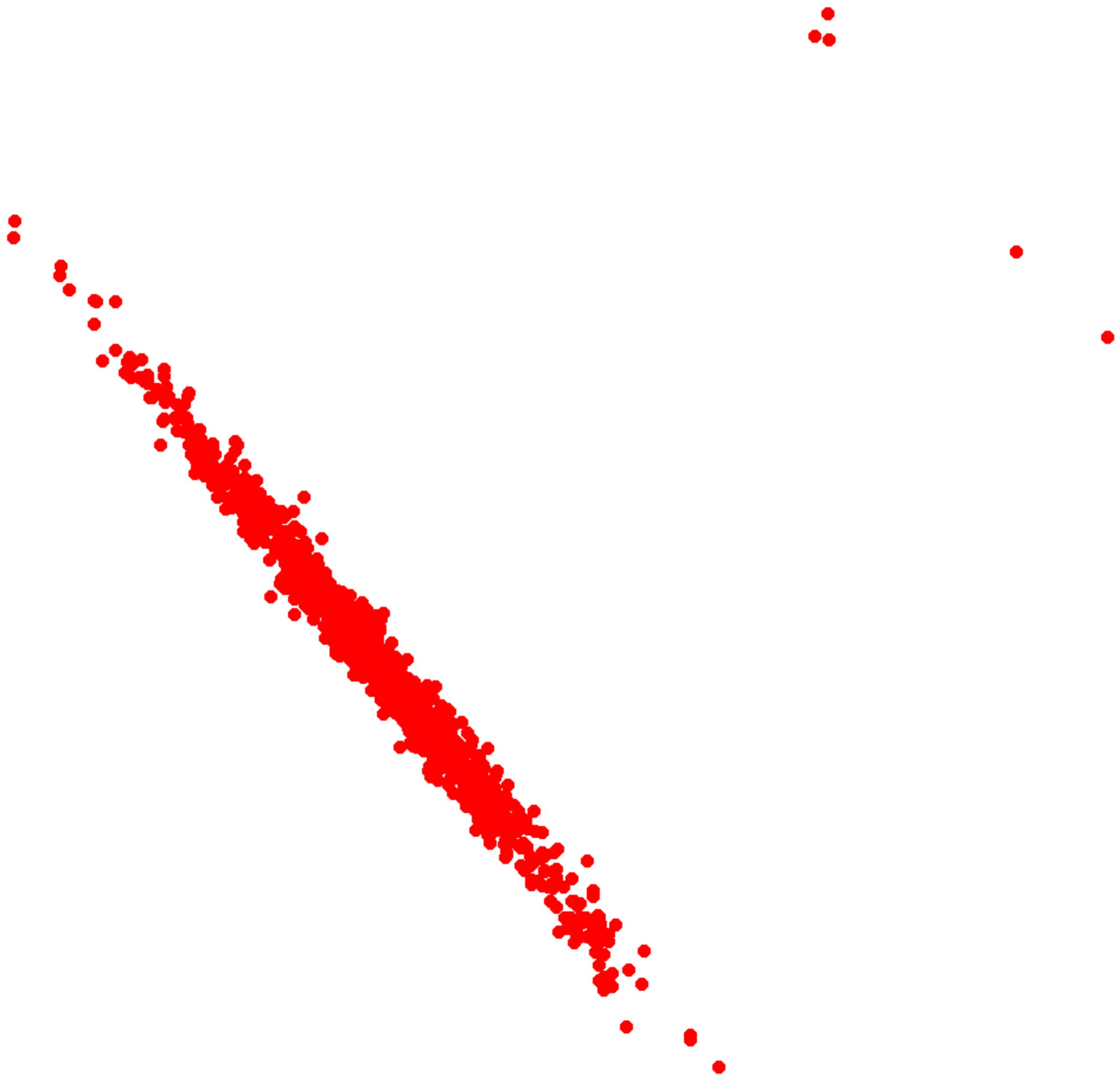
-0.95951



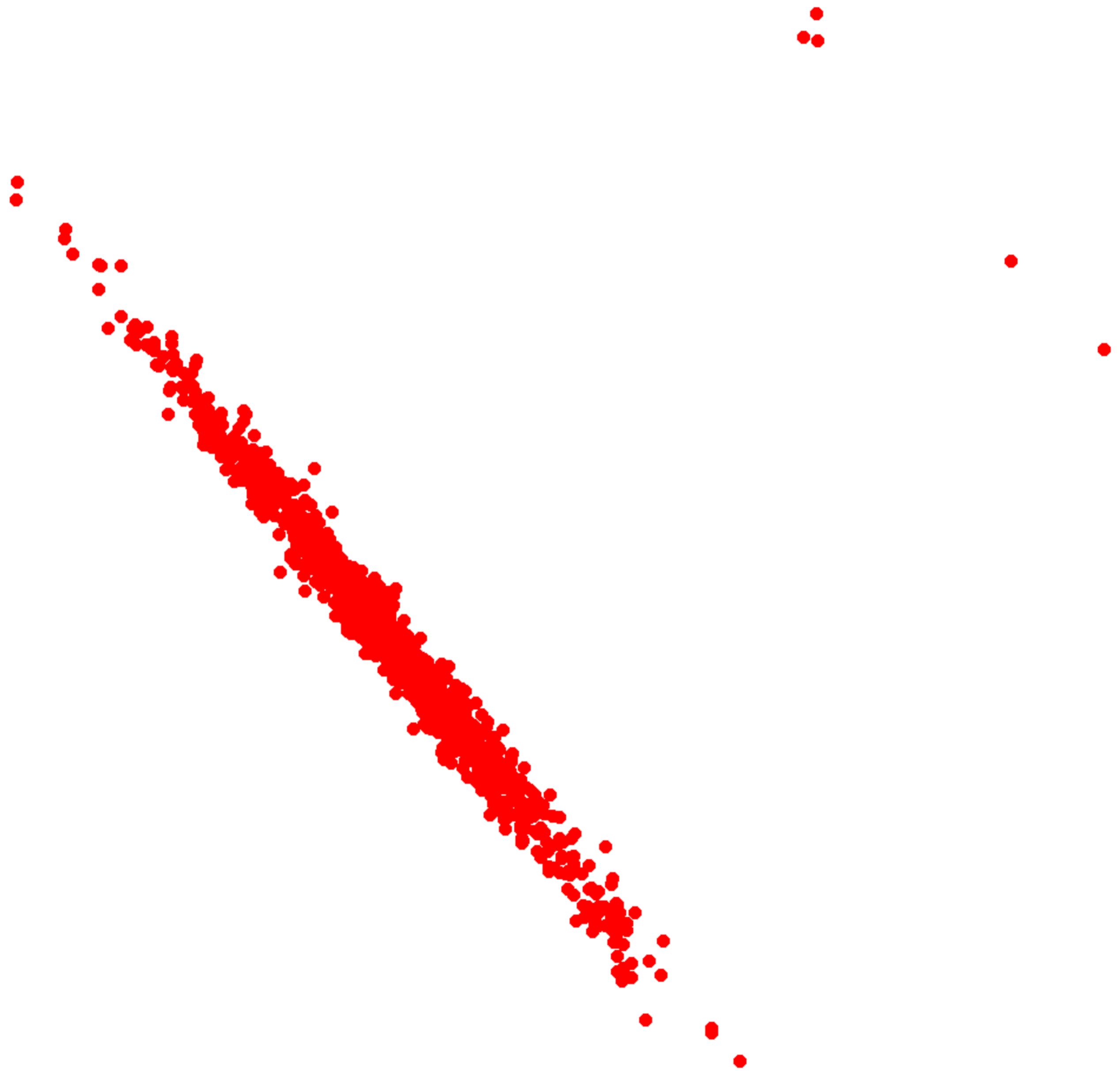
-0.95955



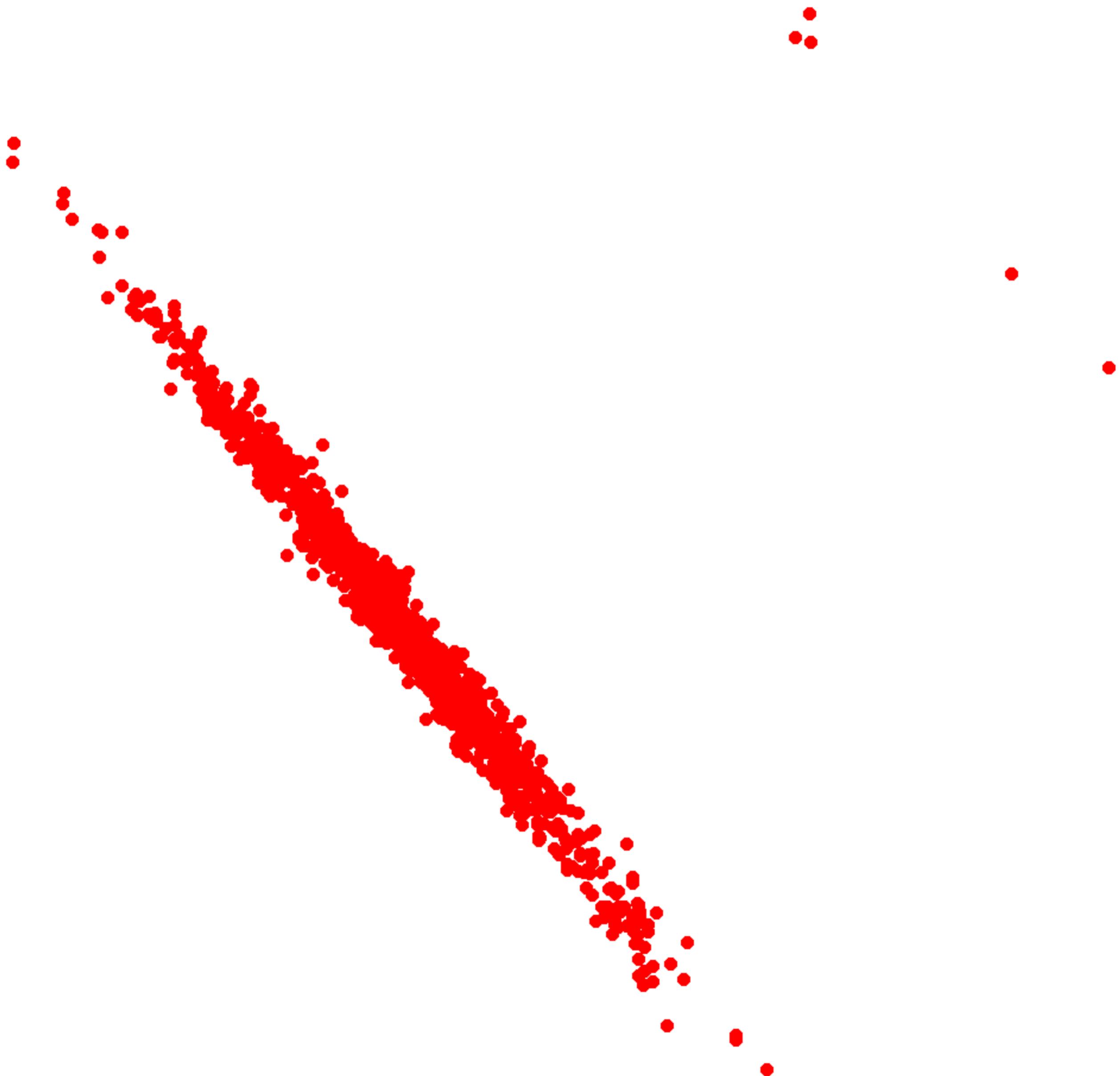
-0.95962



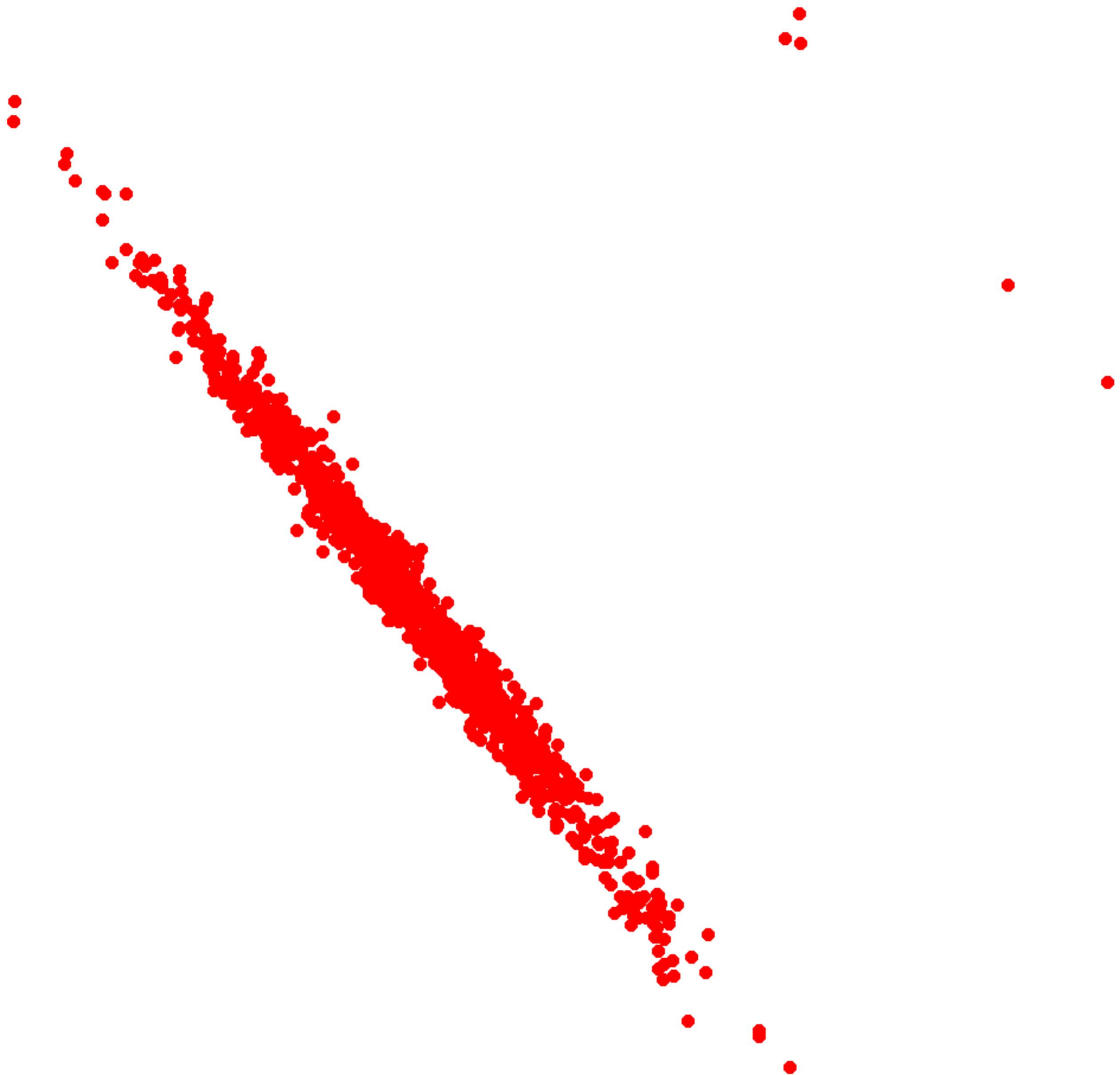
-0.95984



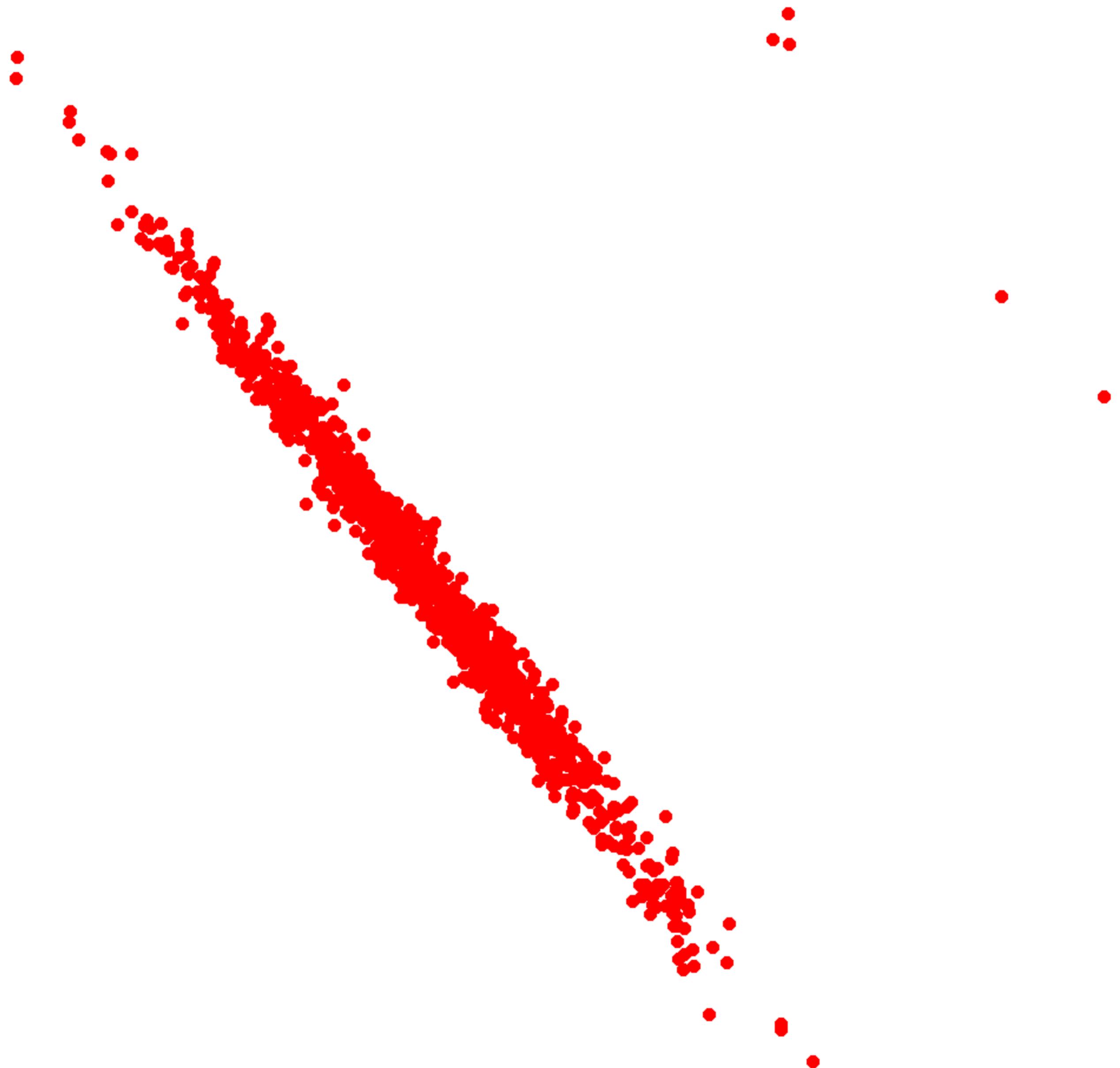
-0.96012



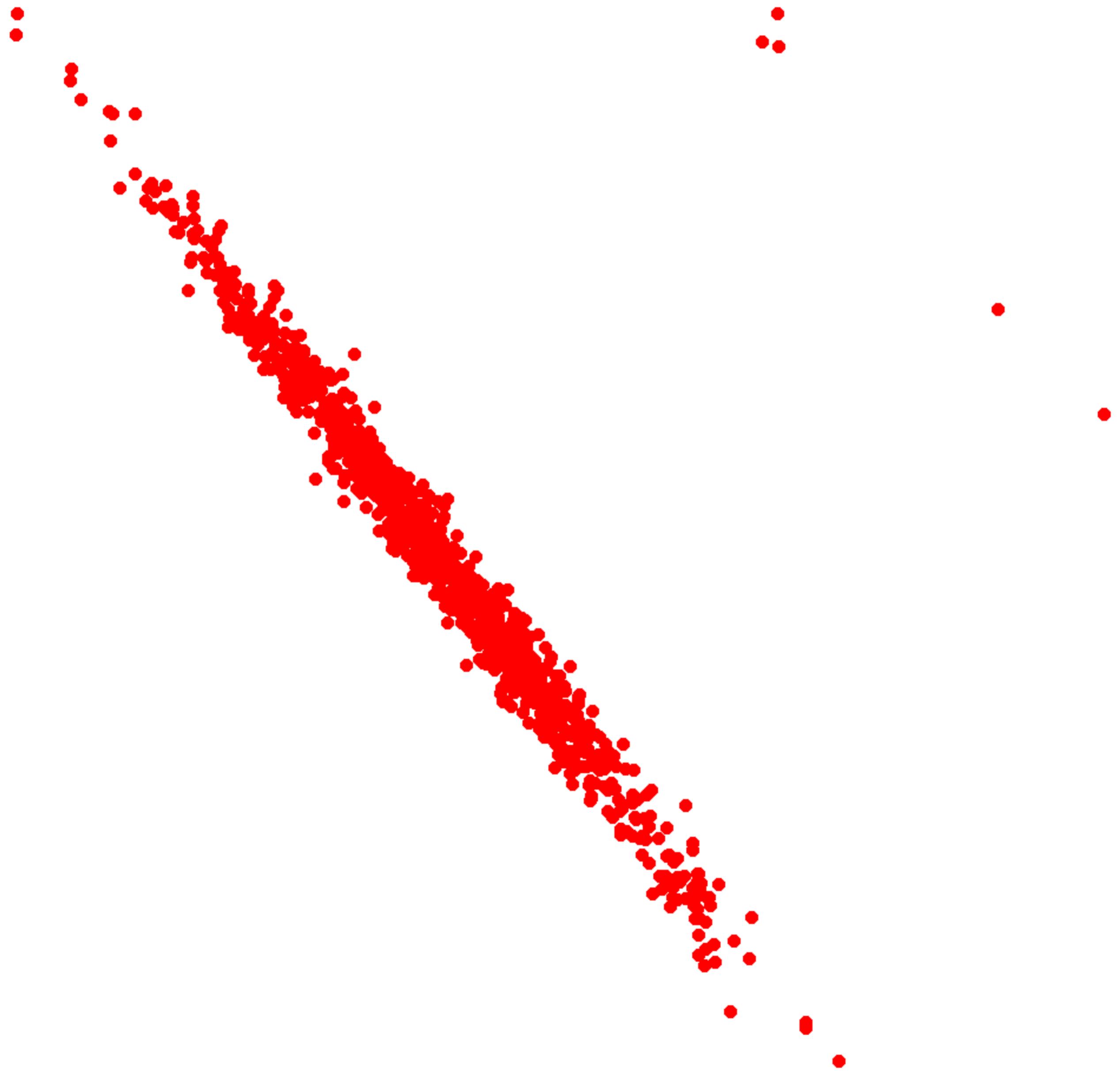
-0.96069



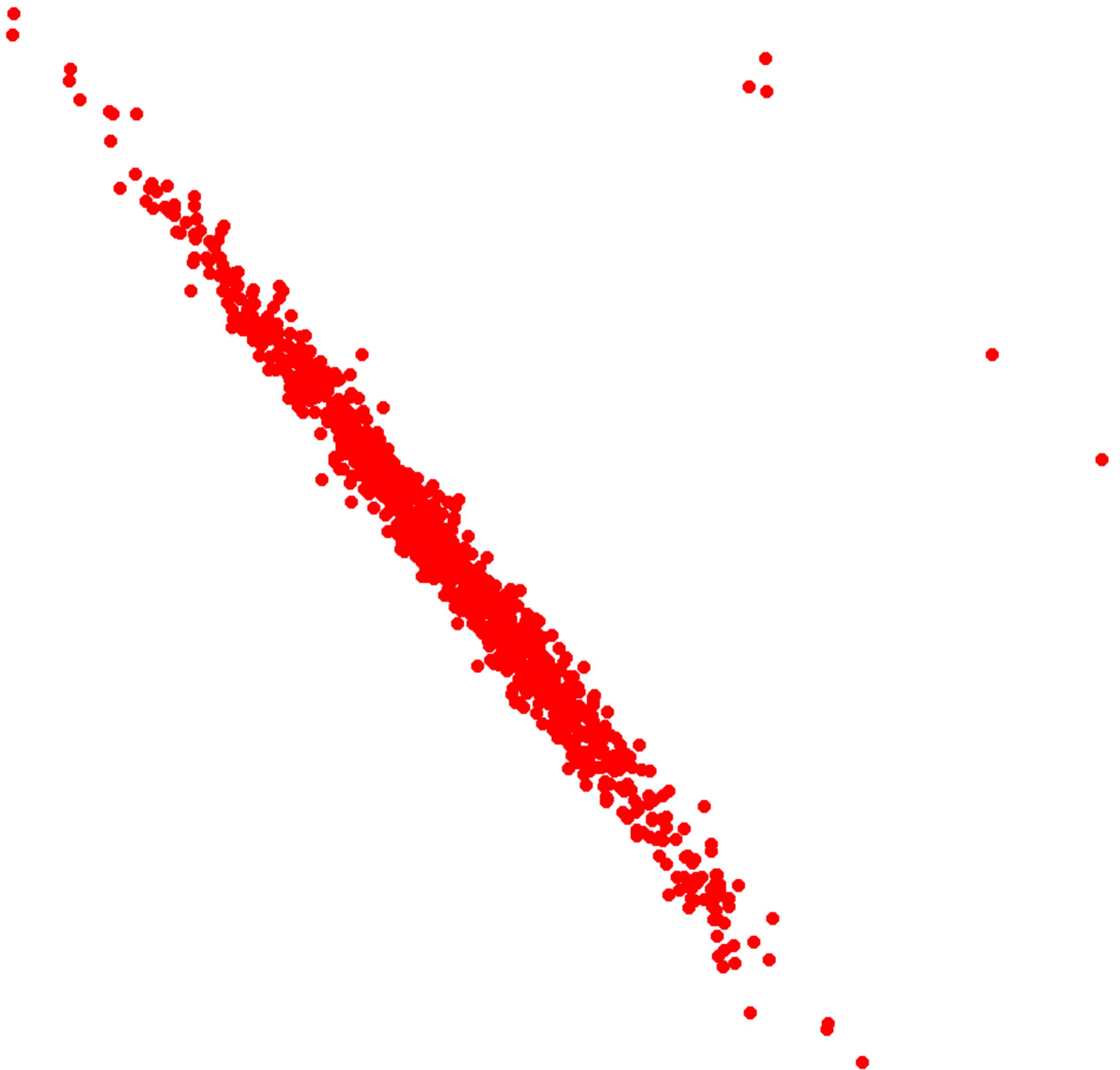
-0.96158



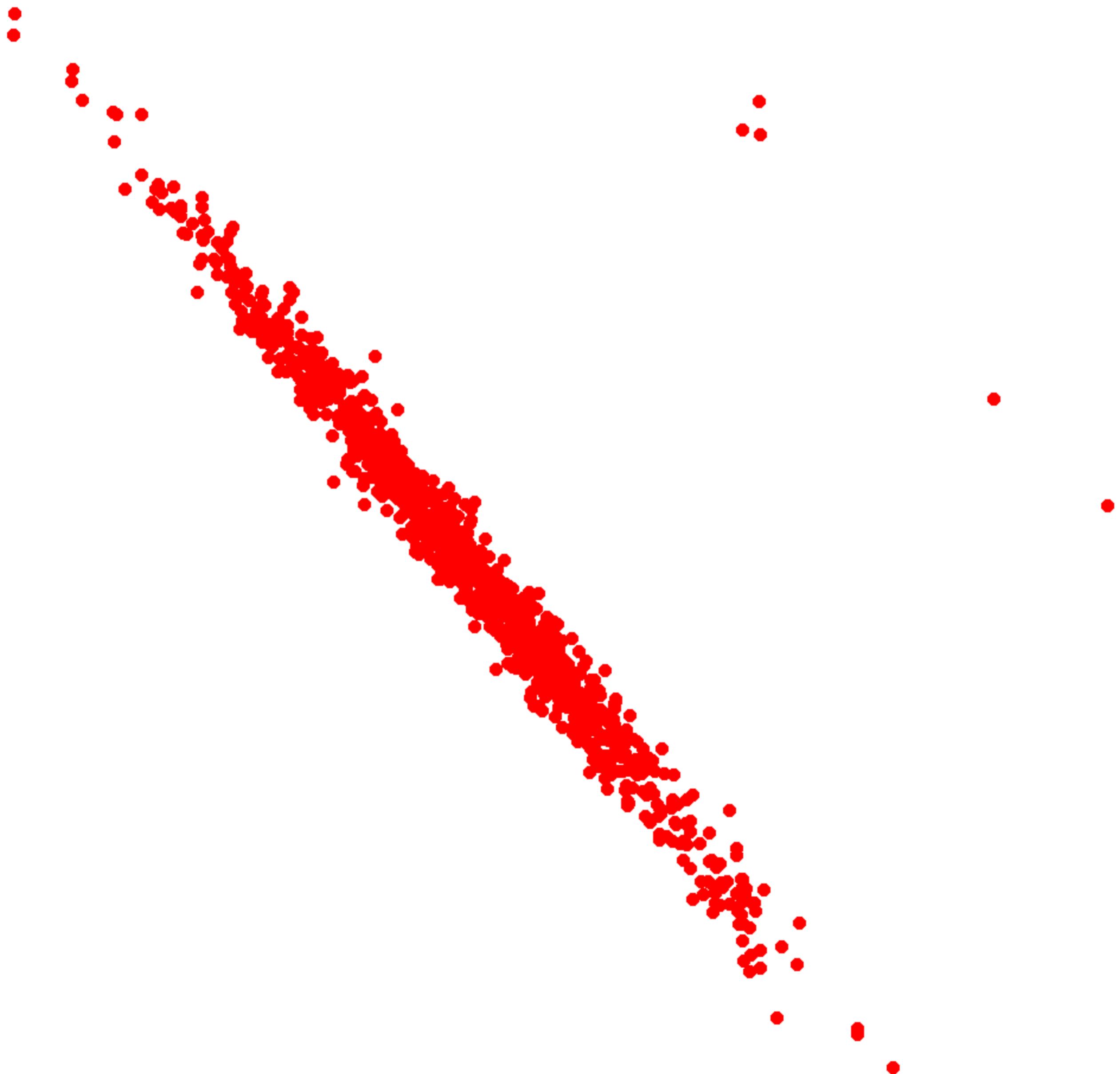
-0.9627



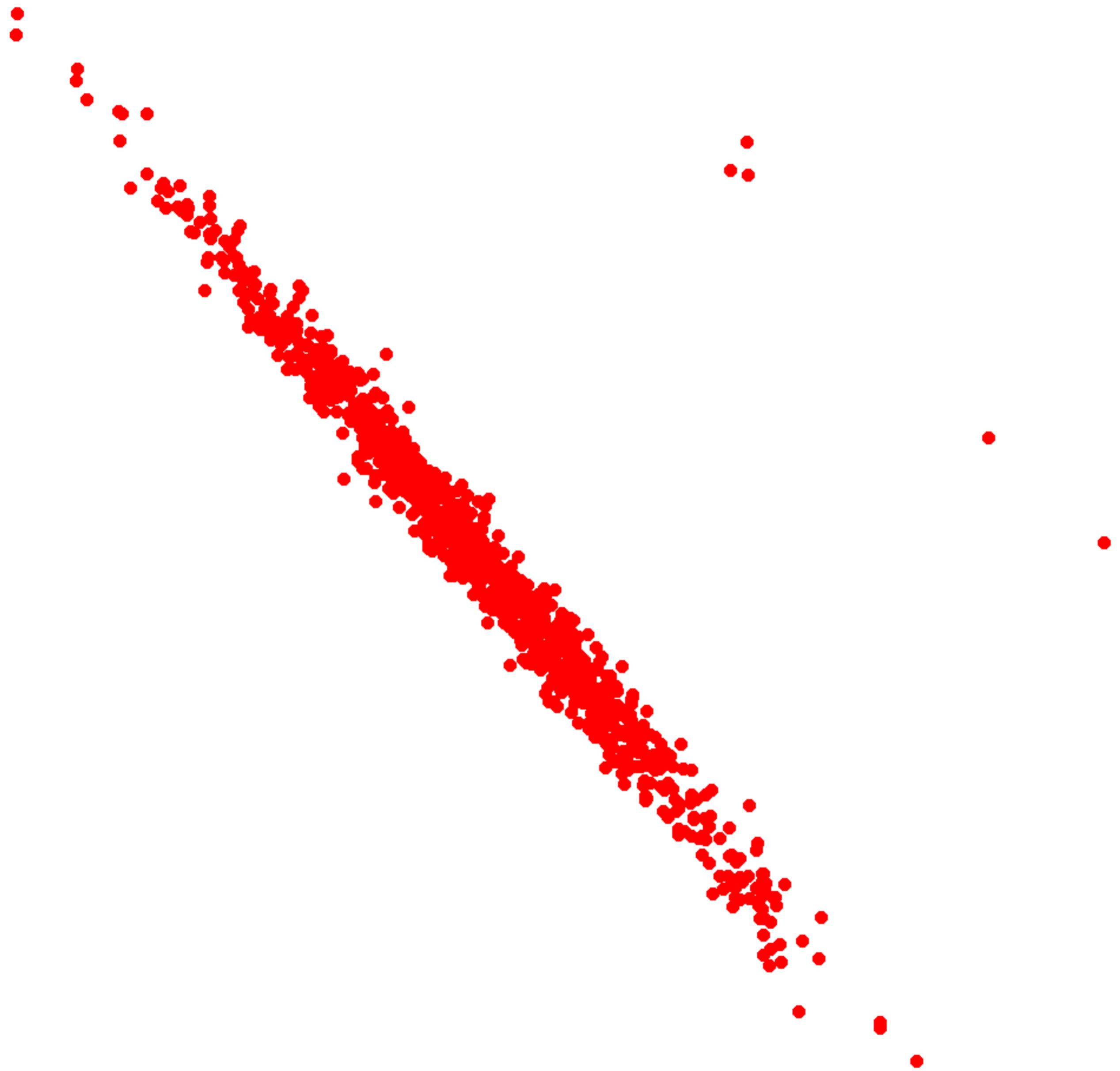
-0.96399



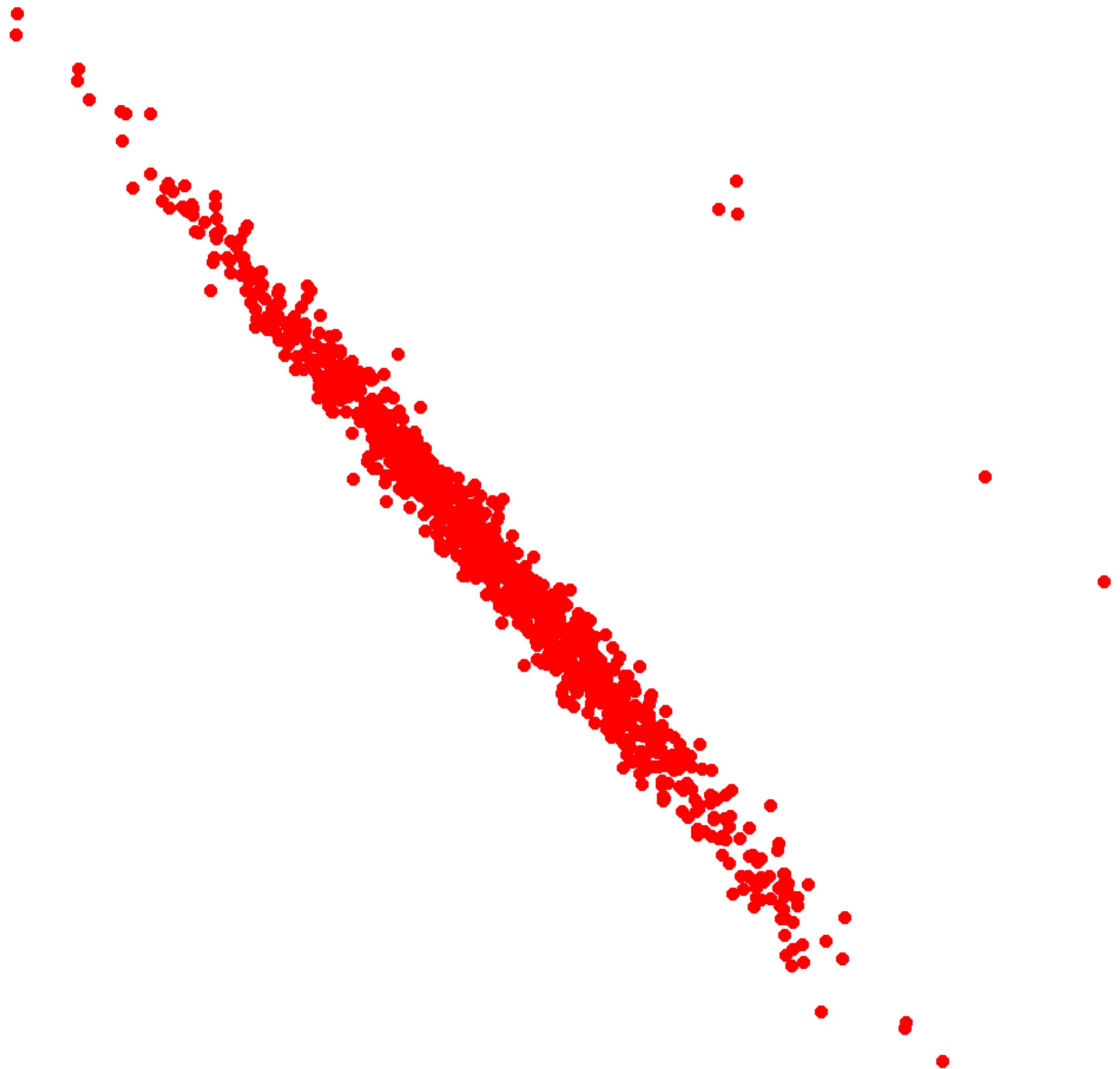
-0.96594



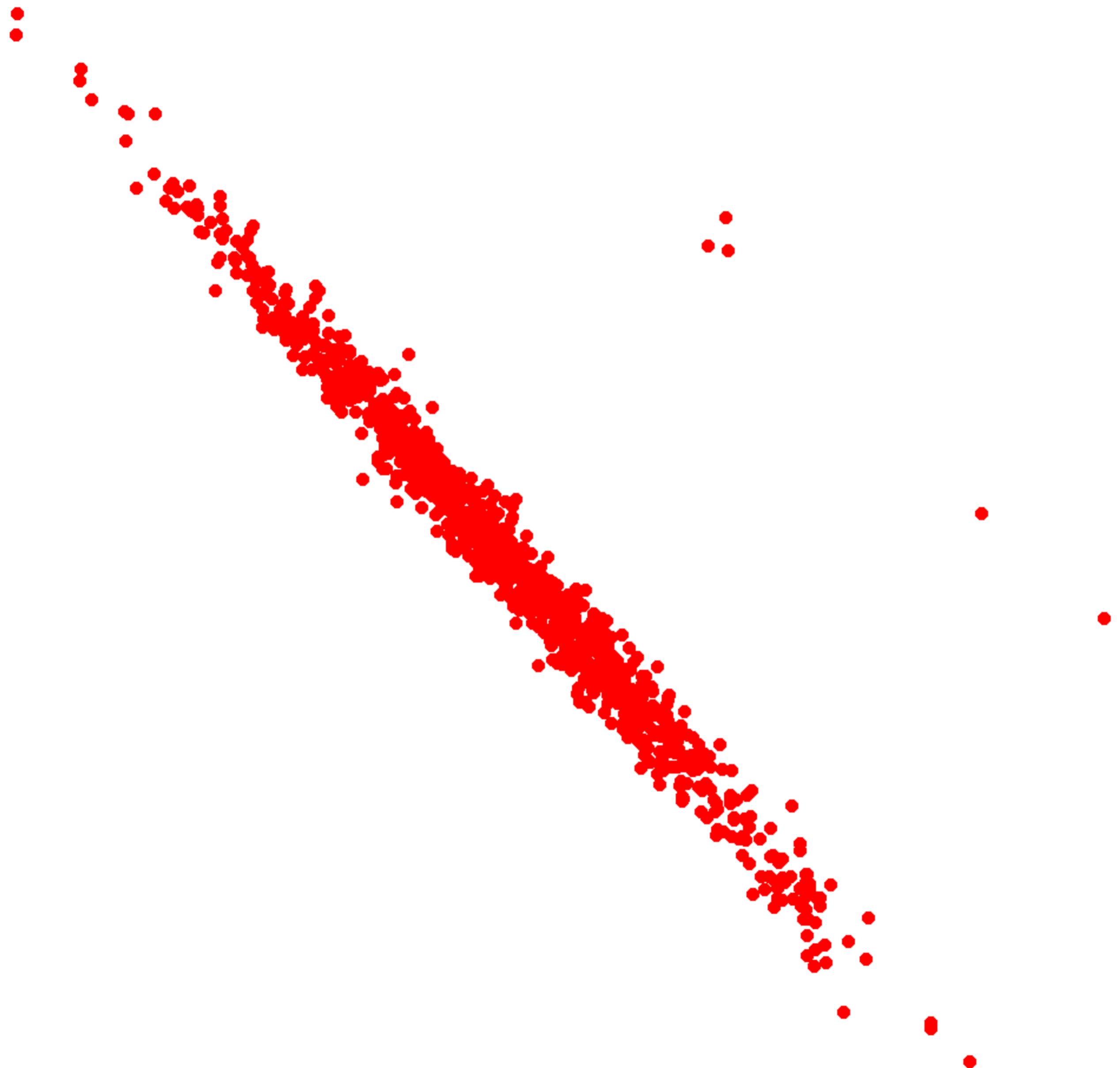
-0.96783



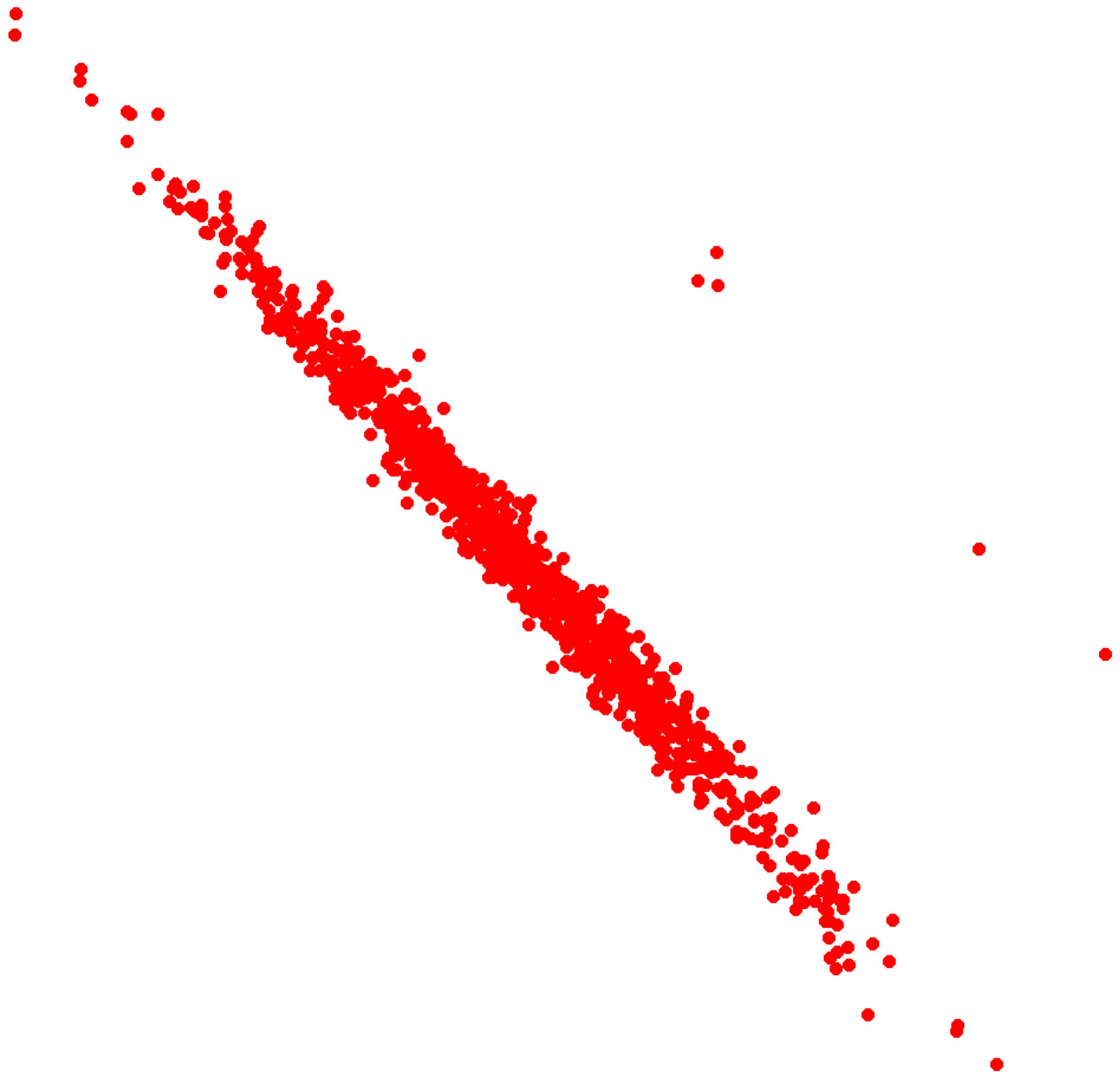
-0.97005



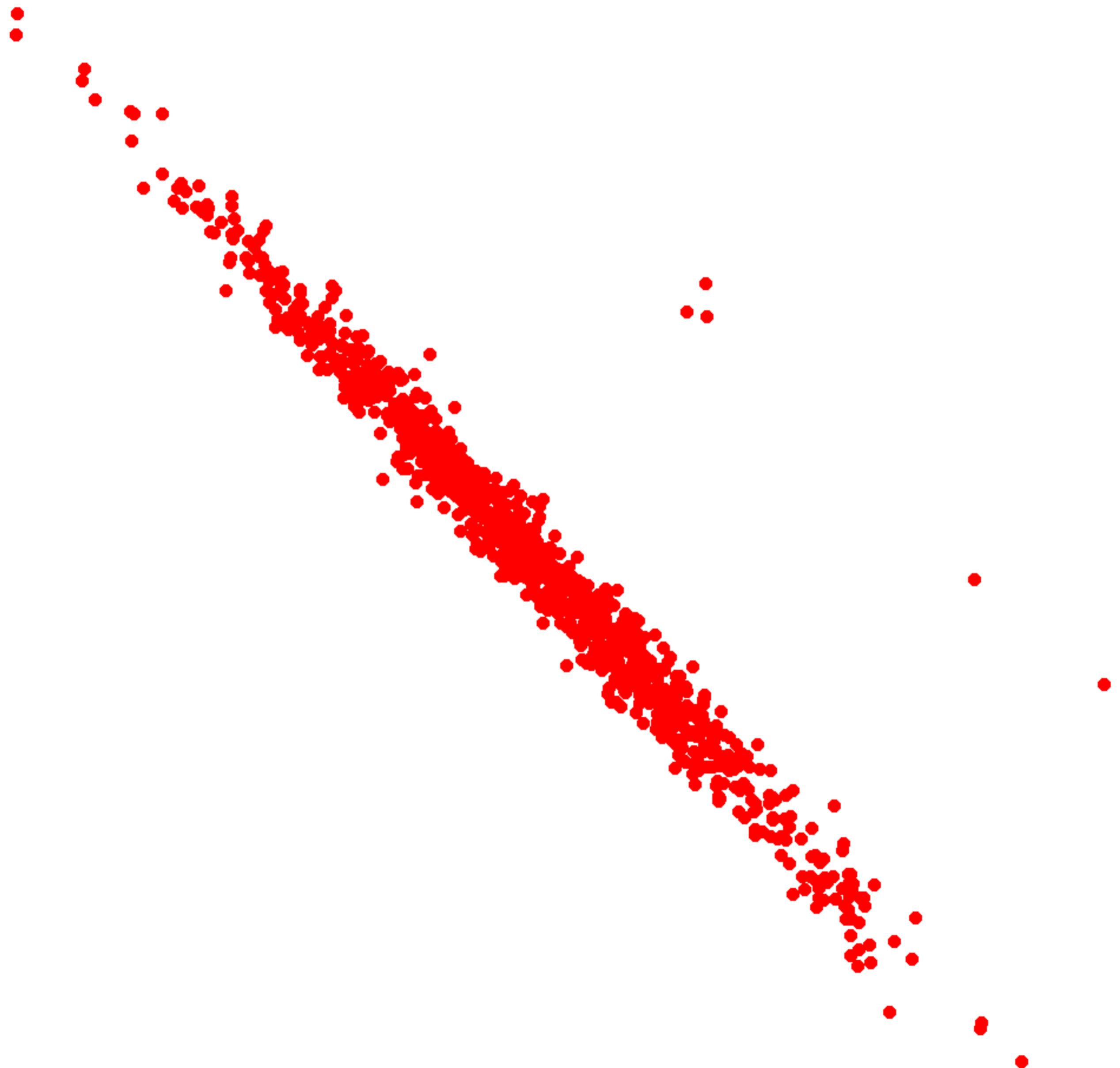
-0.97218



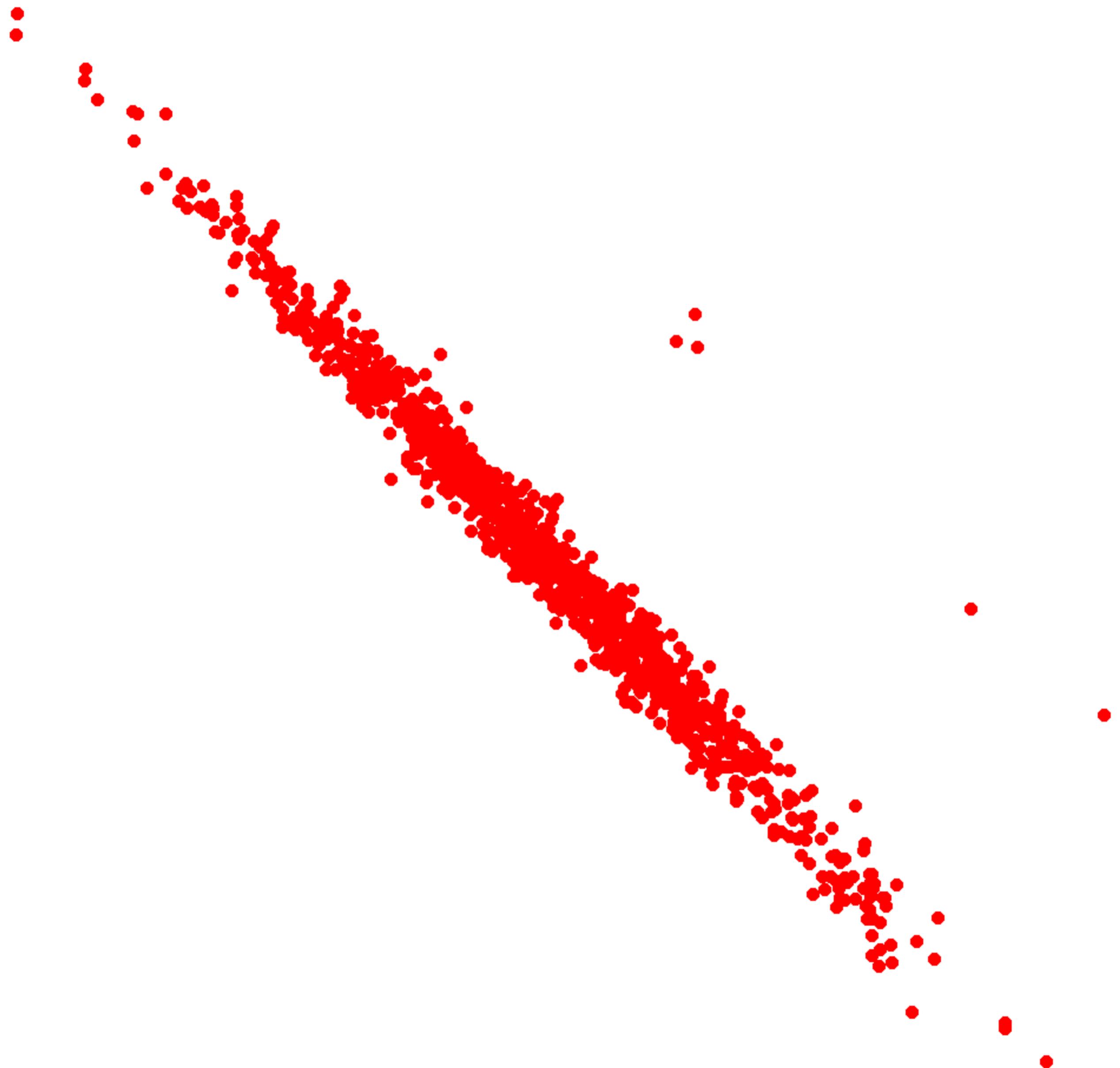
-0.9742



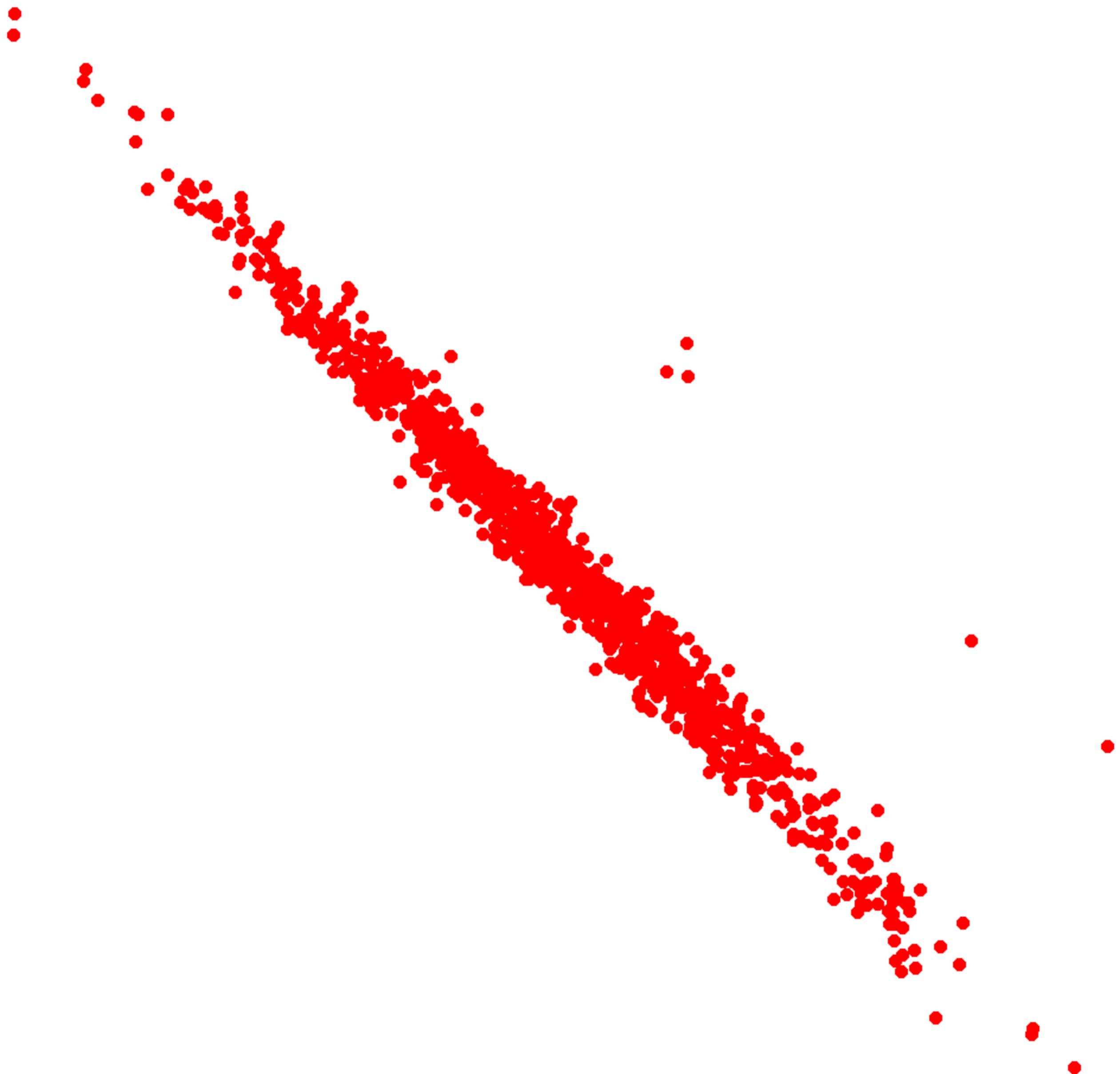
-0.9764



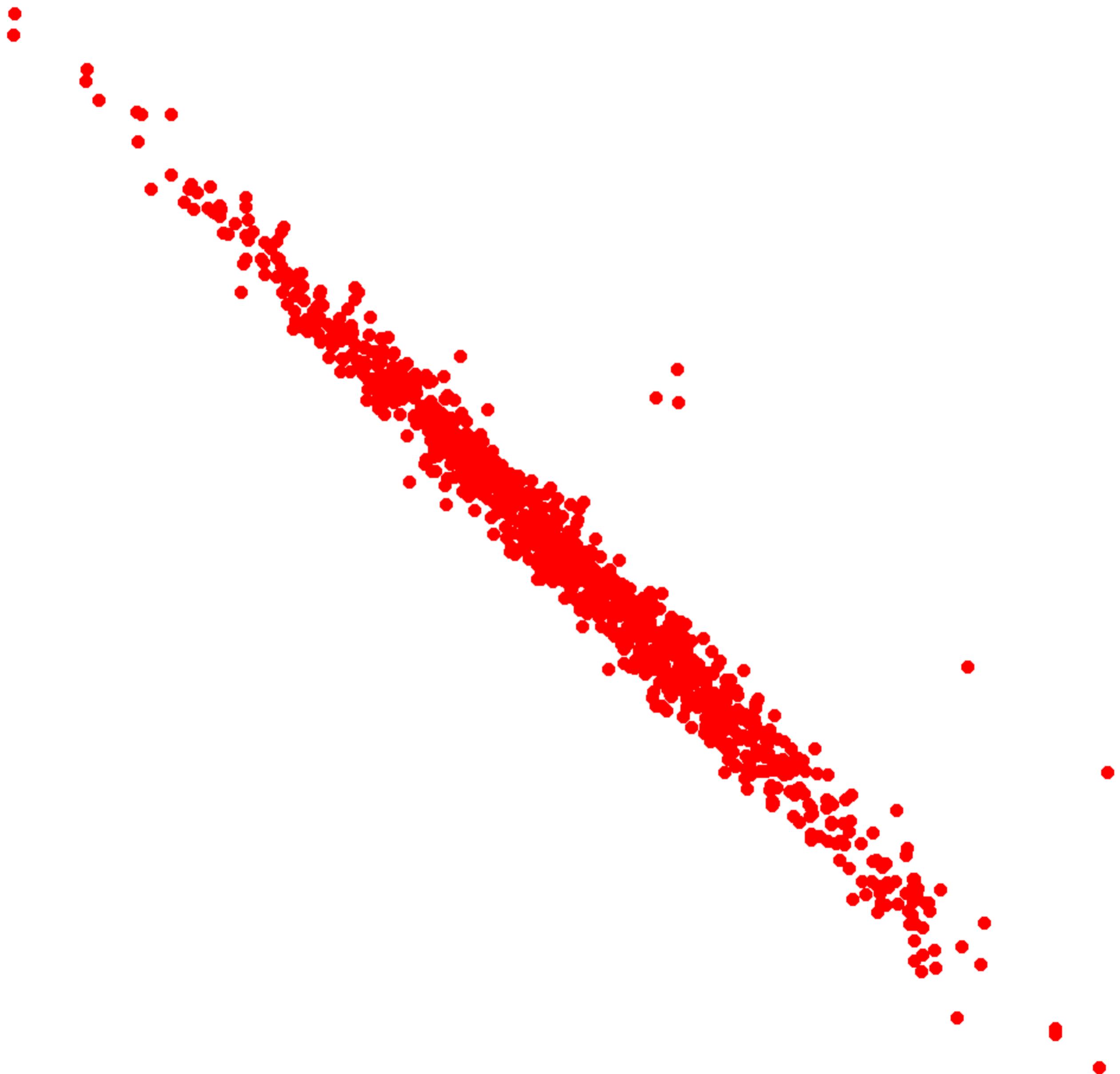
-0.97893



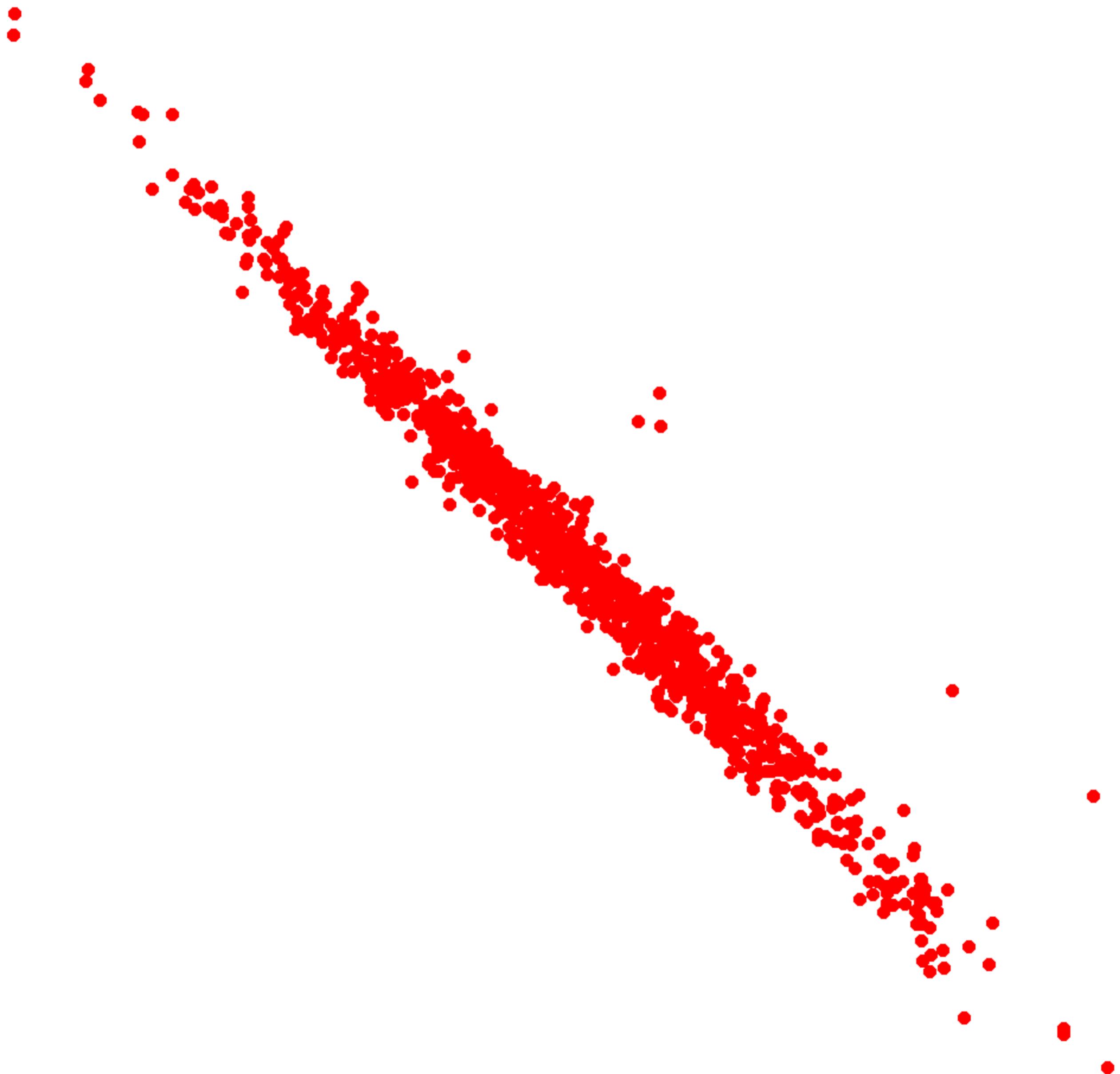
-0.98134



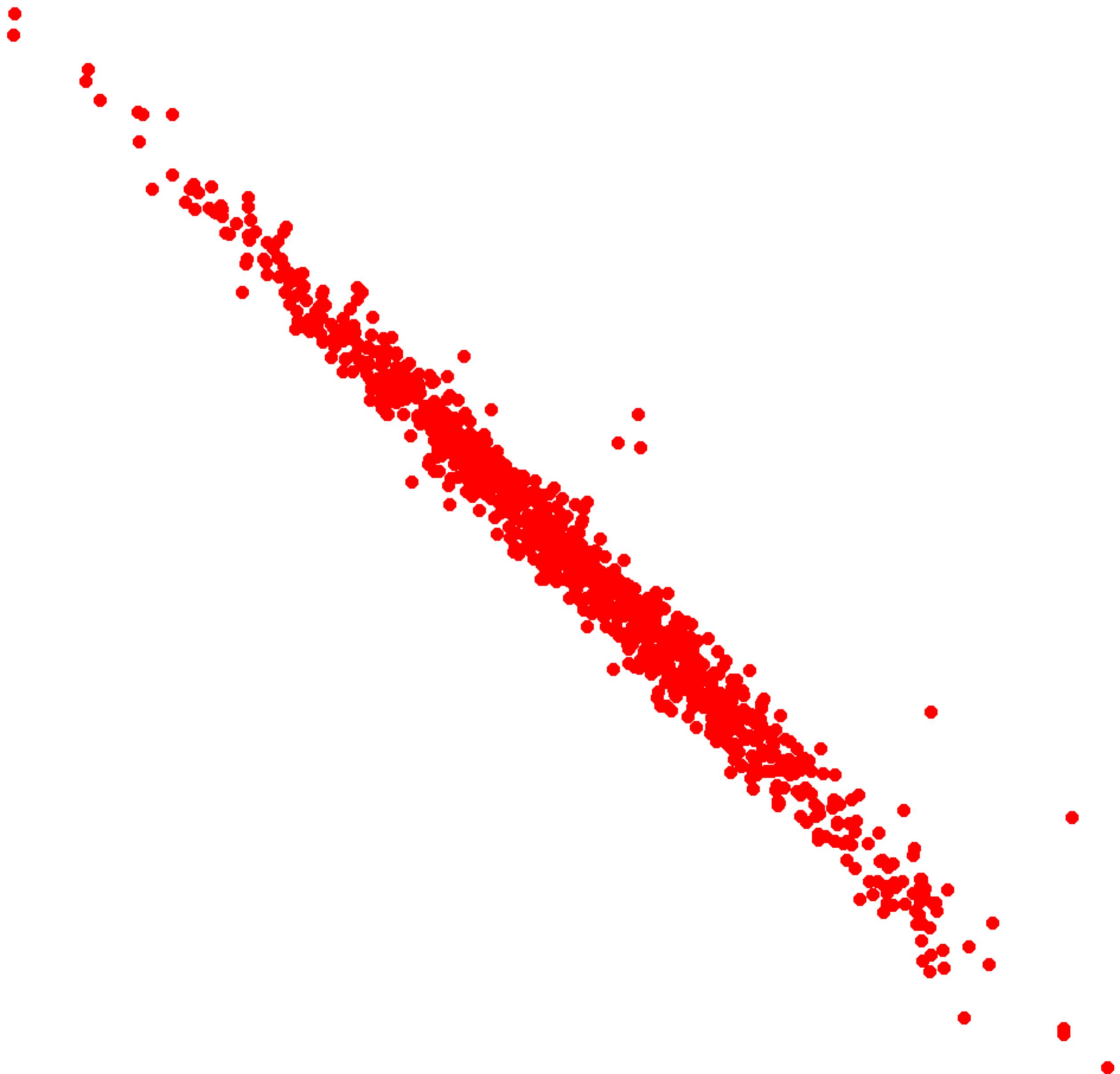
-0.98339



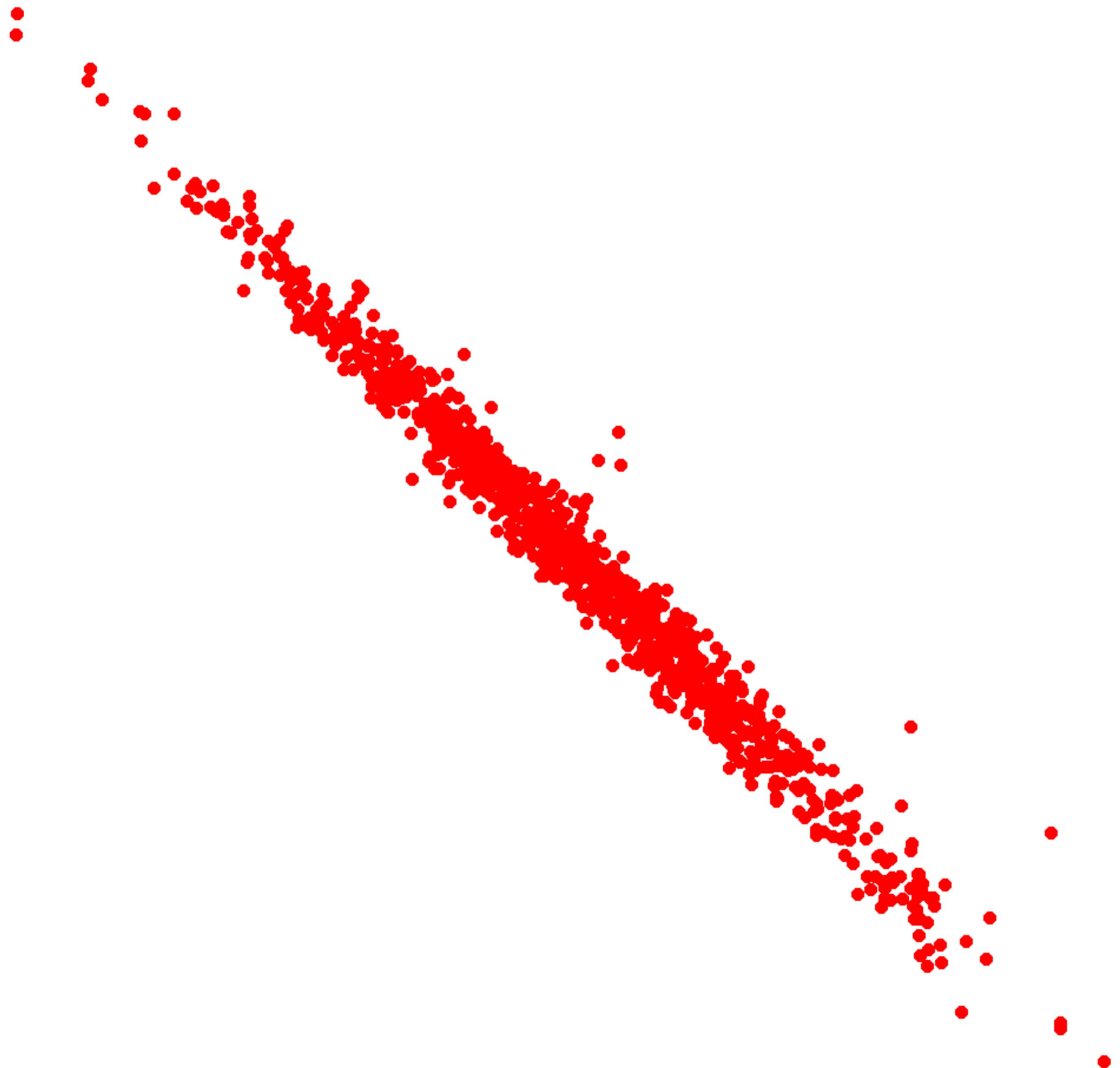
-0.98502



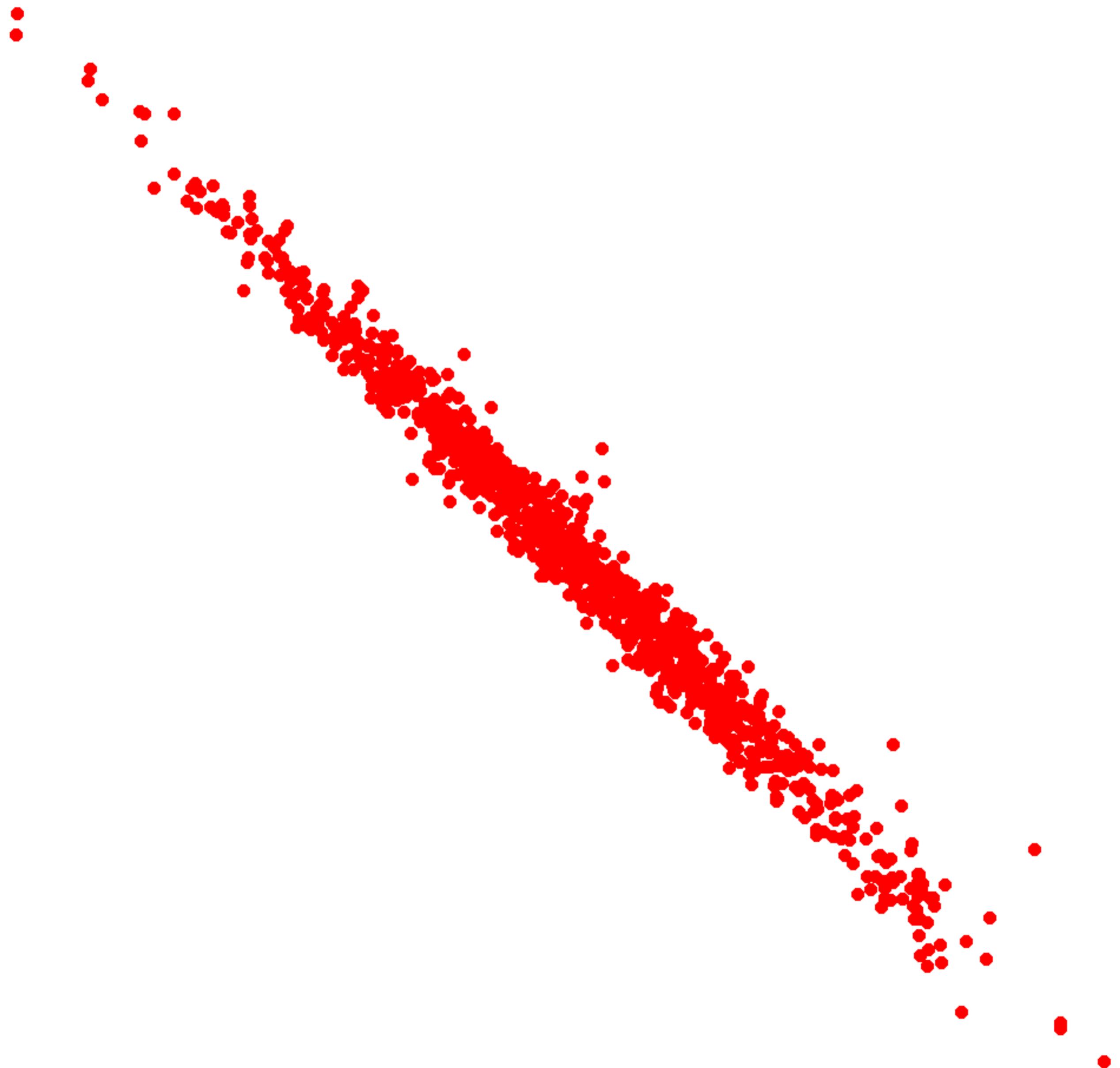
-0.98631



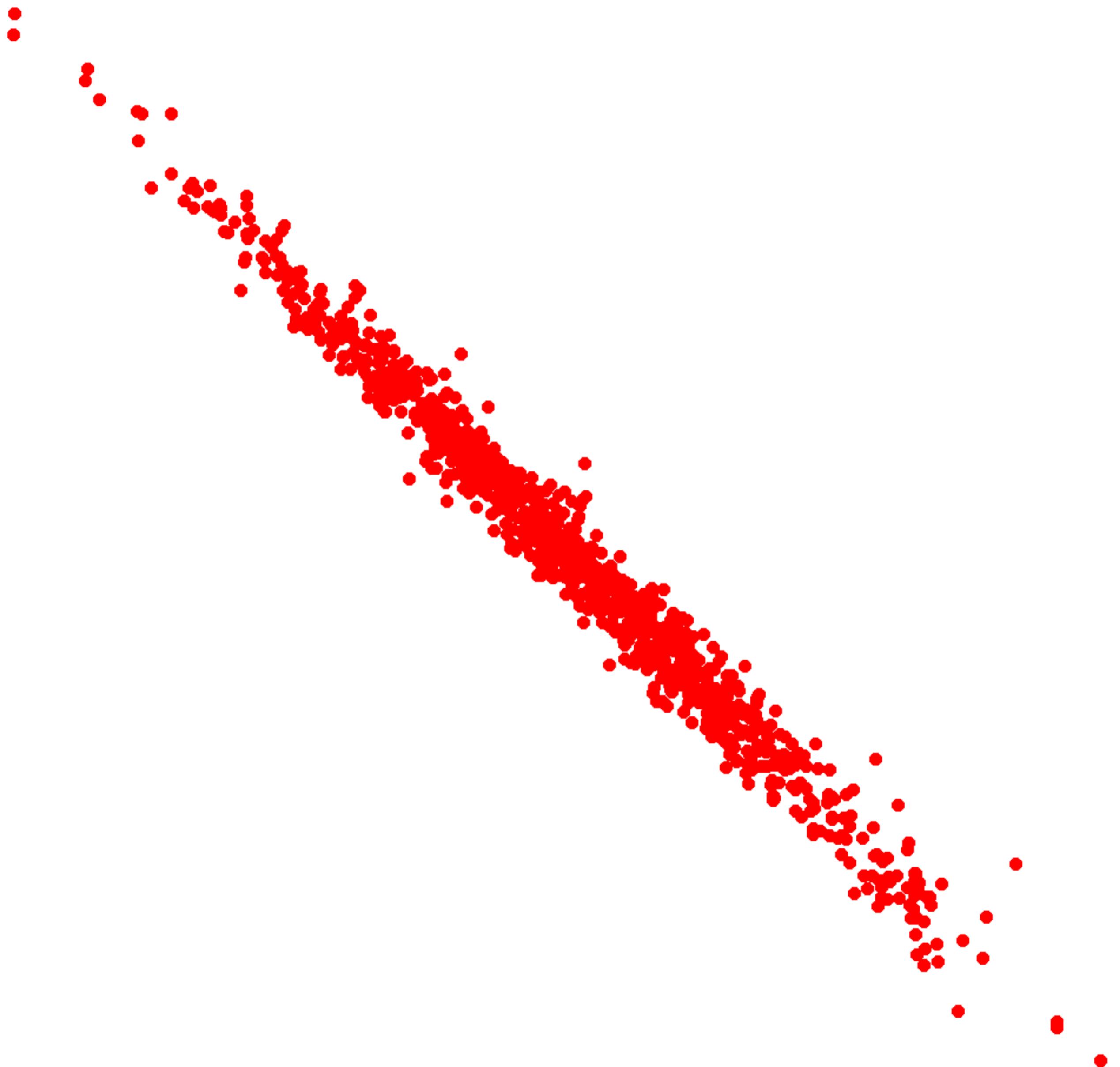
-0.98734



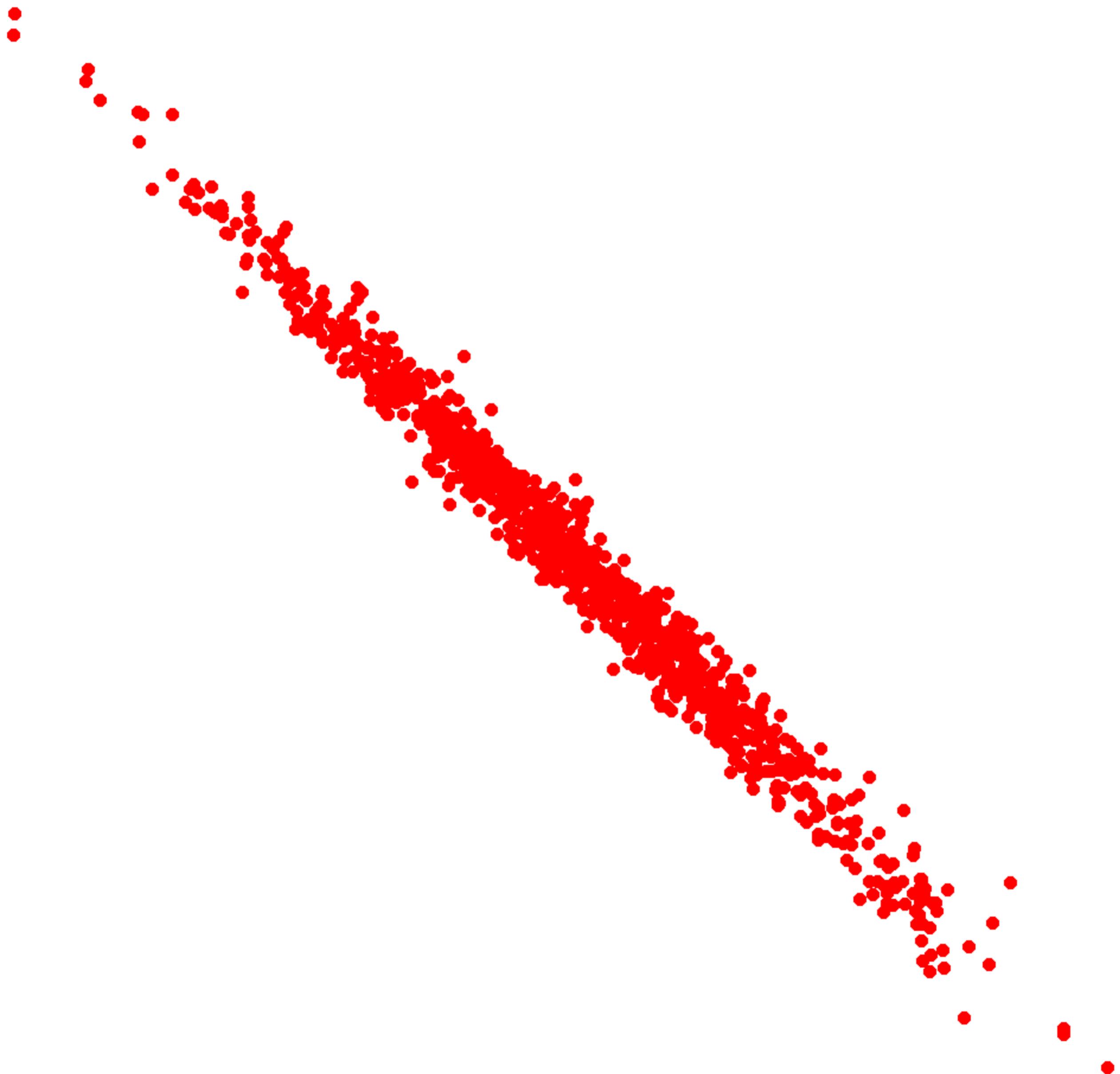
-0.98807



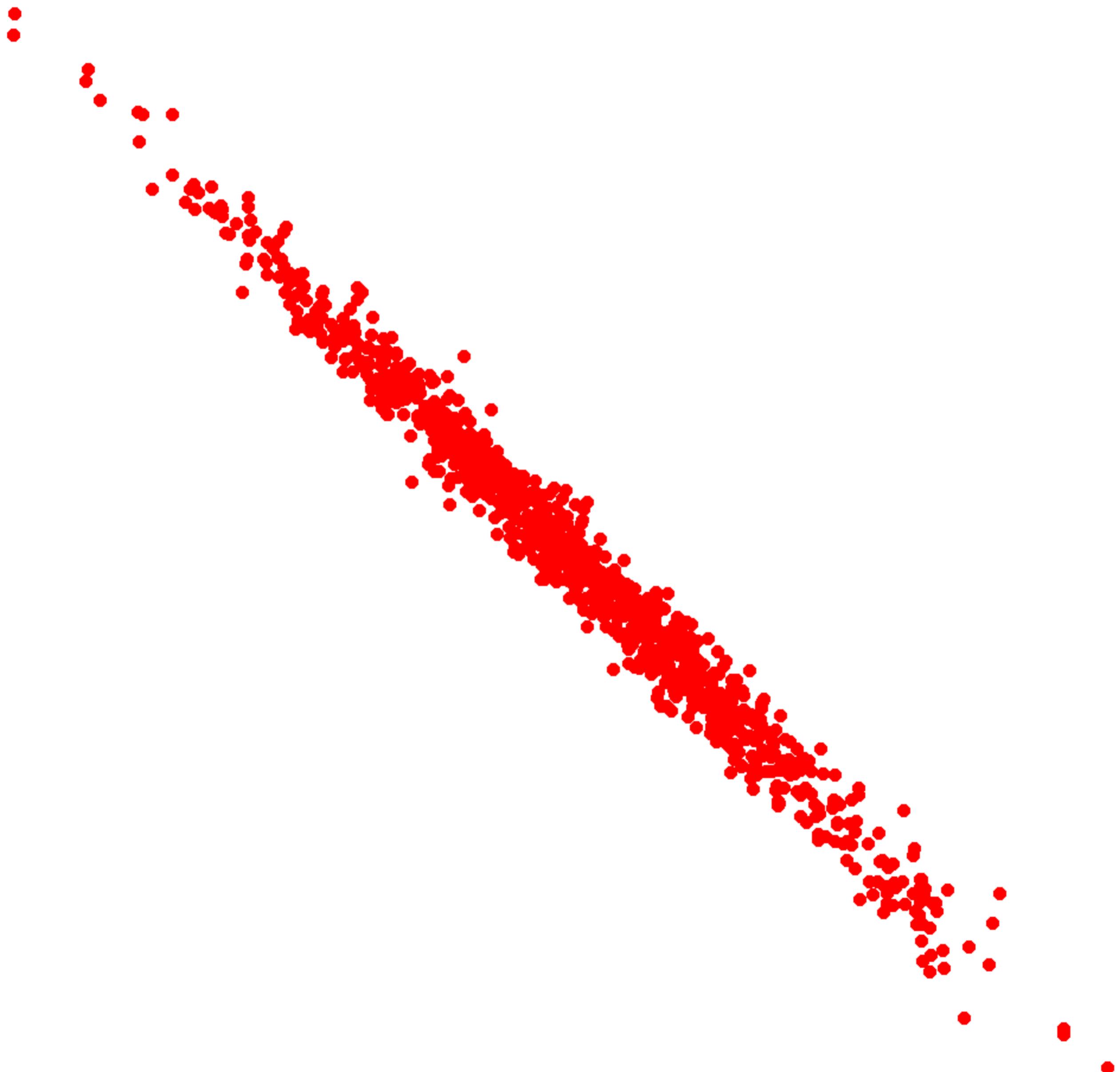
-0.98862



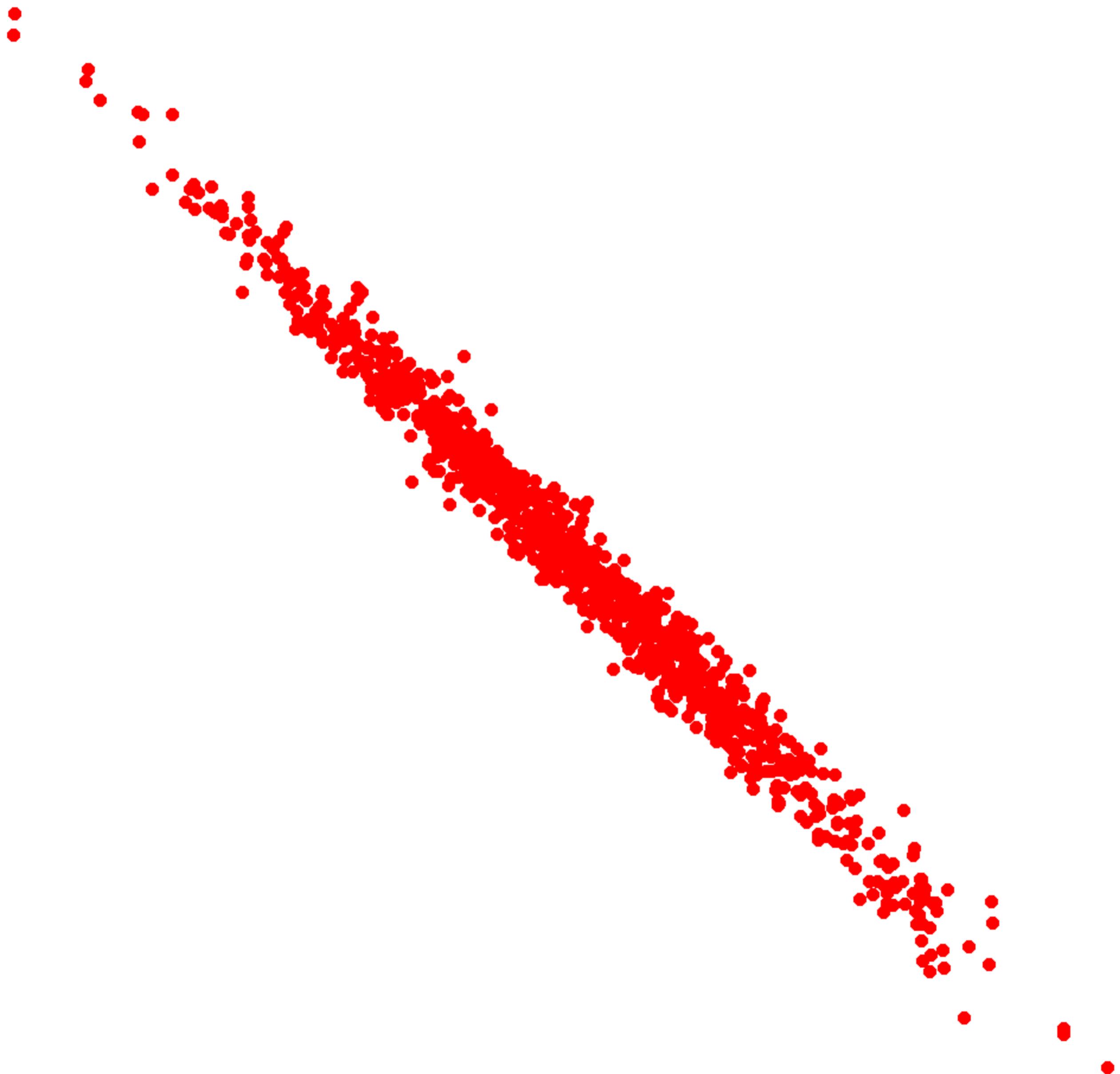
-0.98892



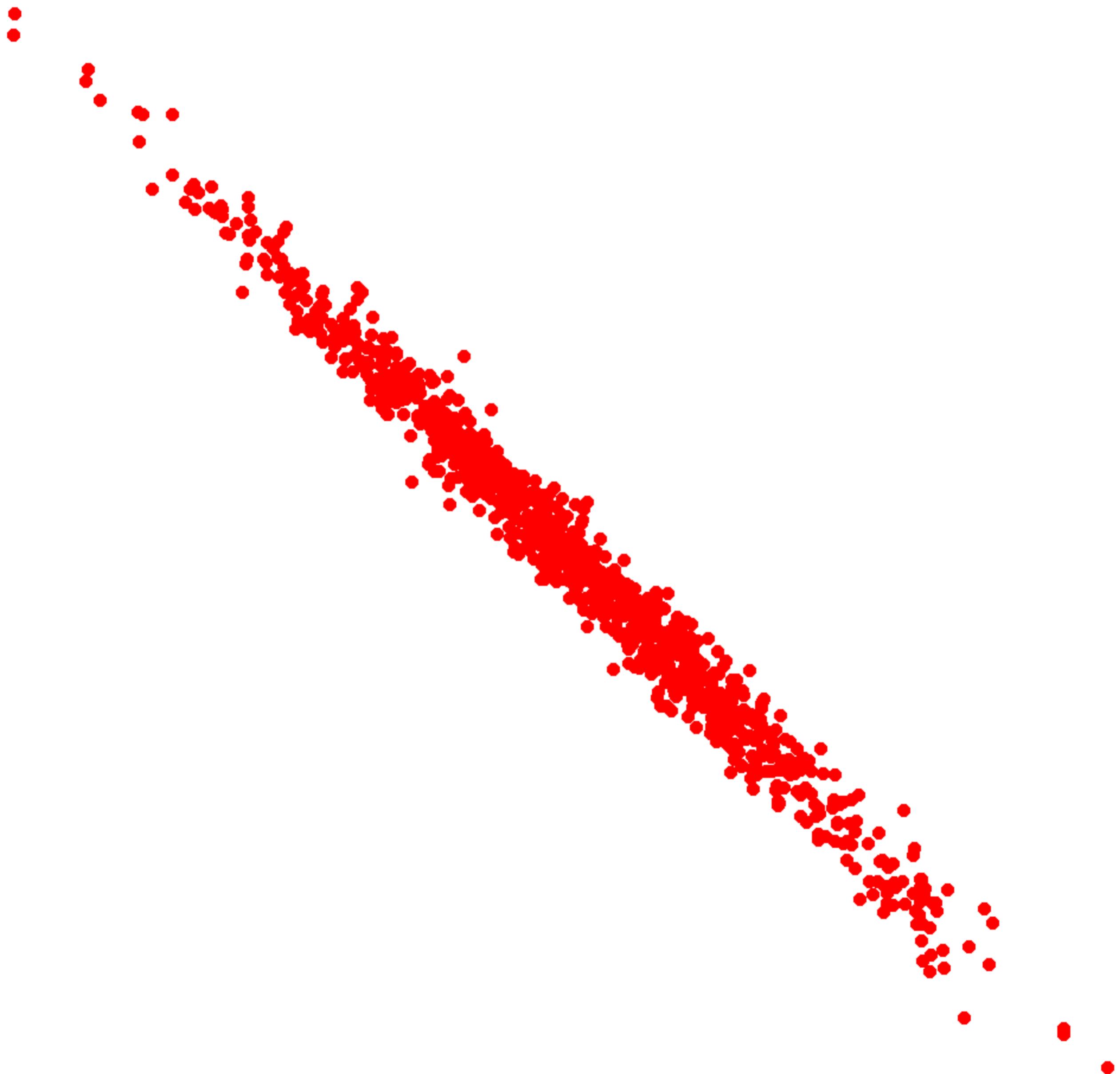
-0.98909



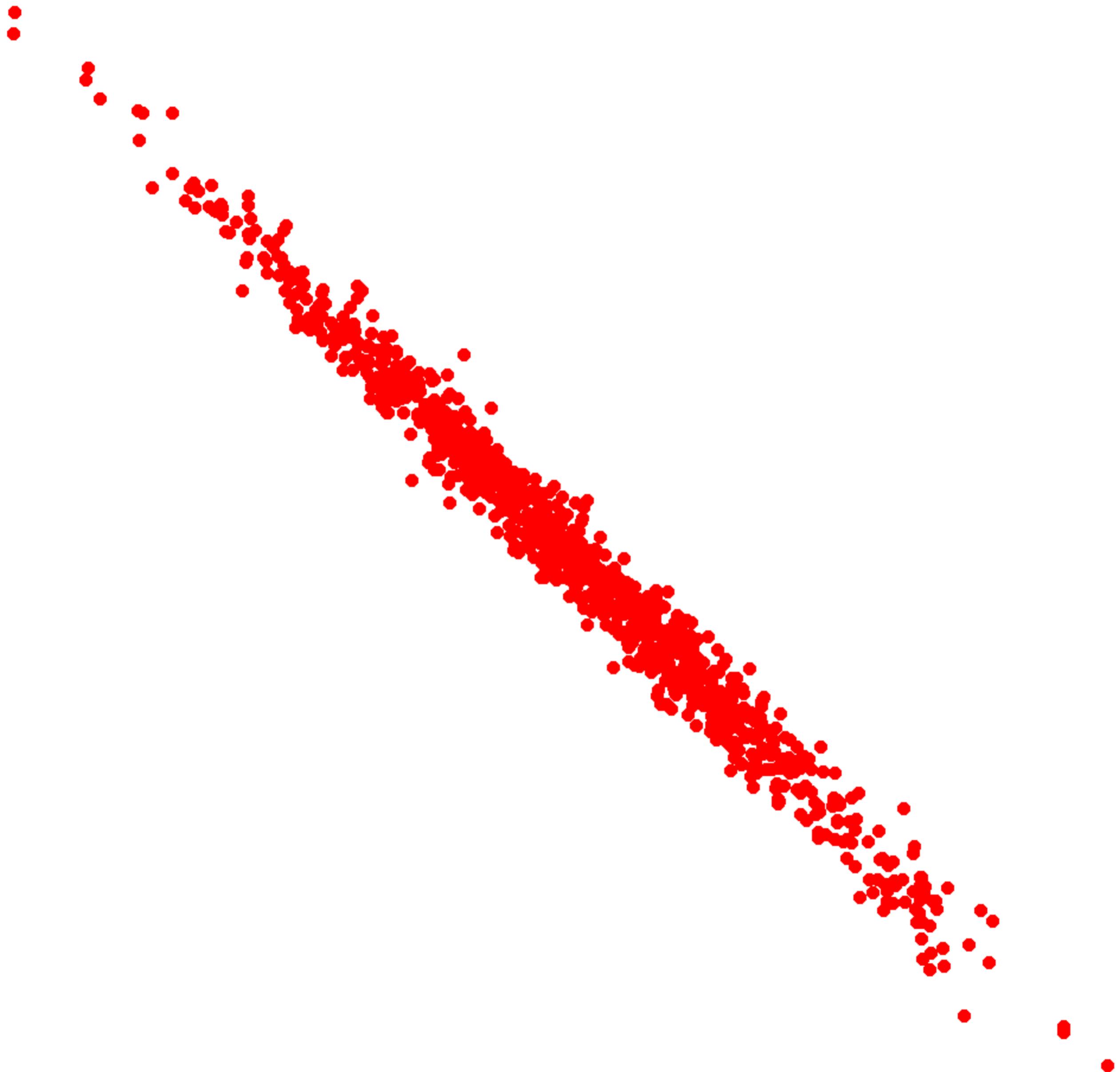
-0.98915



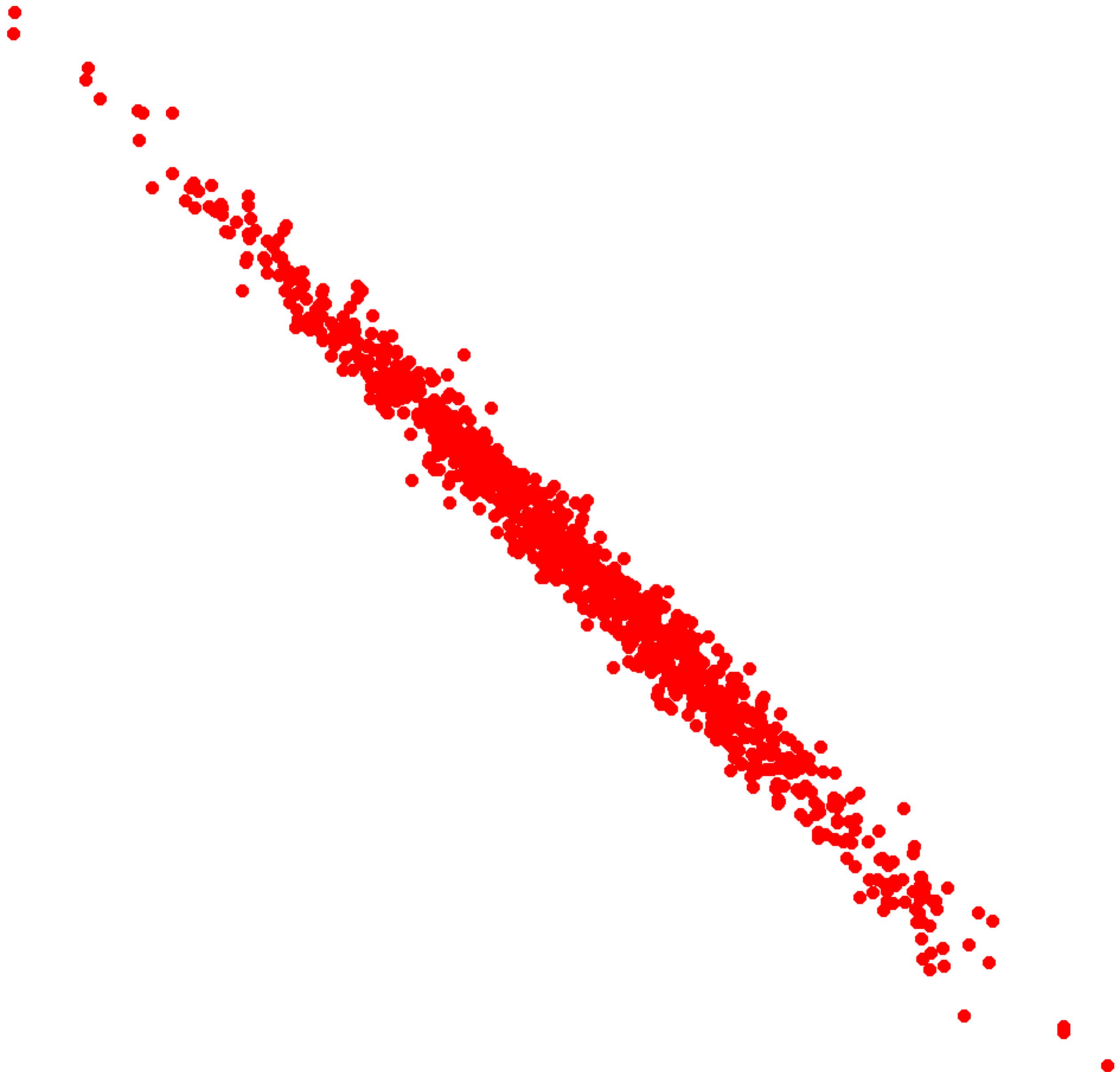
-0.98919



-0.9892



-0.9892



КОРРЕЛЯЦИЯ МЭТЬЮСА И КОЭФФИЦИЕНТ КРАМЕРА

КОРРЕЛЯЦИЯ МЭТЬЮСА

- » Коэффициент корреляции Мэтьюса — мера силы взаимосвязи между двумя бинарными переменными.

КОРРЕЛЯЦИЯ МЭТЬЮСА

- » Коэффициент корреляции Мэтьюса — мера силы взаимосвязи между двумя бинарными переменными.
- » Таблица сопряжённости:

		X_2	
		0	1
X_1	0	a	b
	1	c	d

КОРРЕЛЯЦИЯ МЭТЬЮСА

» Таблица сопряжённости:

		X_2	
		0	1
X_1	0	a	b
	1	c	d

$$\text{MCC}_{X_1 X_2} = \frac{ad - bc}{\sqrt{(a + b)(a + c)(b + d)(c + d)}}$$

КОРРЕЛЯЦИЯ МЭТЬЮСА

$$\text{MCC}_{X_1 X_2} = \frac{ad - bc}{\sqrt{(a + b)(a + c)(b + d)(c + d)}}$$

- › $\text{MCC}_{X_1 X_2} \in [-1, 1]$
- ▶ 0 — полное отсутствие взаимосвязи,
- ▶ 1 $\Leftrightarrow b = c = 0$
- ▶ $-1 \Leftrightarrow a = d = 0$

ТАБЛИЦА СОПРЯЖЁННОСТИ $K_1 \times K_2$

X_1	X_2	1	\dots	j	\dots	K_2
1						
\vdots						
i				n_{ij}		
\vdots						
K_1						

- » Коэффициент V Крамера — мера силы взаимосвязи между двумя категориальными переменными.

$$\phi_c(X_1^n, X_2^n) = \sqrt{\frac{\chi^2(X_1^n, X_2^n)}{n(\min(K_1, K_2) - 1)}}$$

$$\phi_c(X_1^n, X_2^n) = \sqrt{\frac{\chi^2(X_1^n, X_2^n)}{n(\min(K_1, K_2) - 1)}}$$

» $\phi_c(X_1^n, X_2^n) \in [0, 1]$

- ▶ 0 — полное отсутствие взаимосвязи,
- ▶ 1 — совпадение переменных (с точностью до переименования уровней).

ПАРЫ ПЕРЕМЕННЫХ РАЗНЫХ ТИПОВ

- » Между категориальными и непрерывными признаками корреляции считать не нужно!
- » Пусть $X_1 \in \mathbb{R}$, $X_2 \in \{0, 1\}$;
 X_1 и X_2 положительно коррелированы, если
 $\mathbb{E}(X_1 | X_2 = 1) > \mathbb{E}(X_1 | X_2 = 0)$.

ПАРЫ ПЕРЕМЕННЫХ РАЗНЫХ ТИПОВ

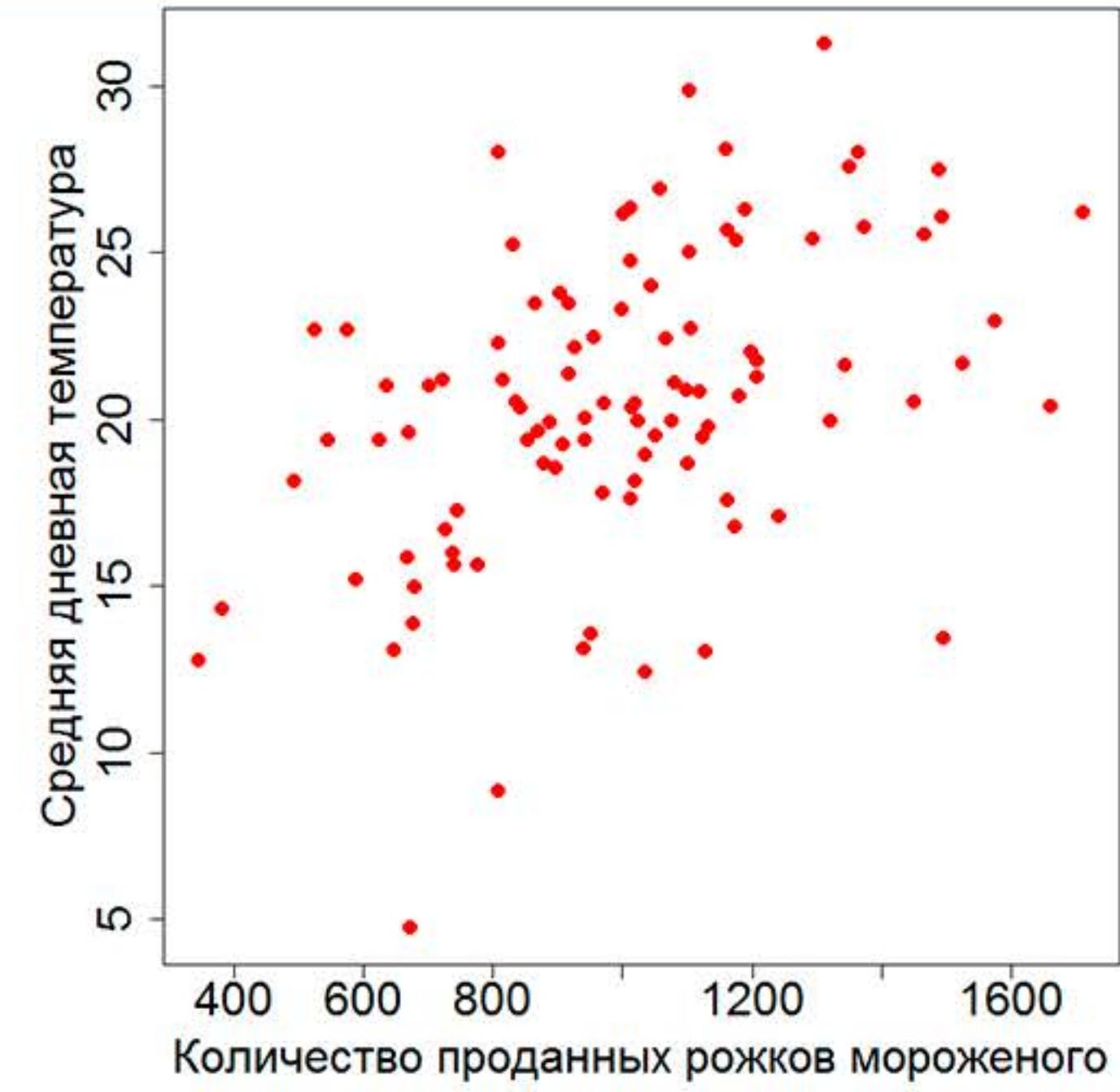
- » Пусть $X_1 \in \mathbb{R}$, $X_2 \in \{0, 1\}$;
 X_1 и X_2 положительно коррелированы, если
 $\mathbb{E}(X_1 | X_2 = 1) > \mathbb{E}(X_1 | X_2 = 0)$.
- » Мера взаимосвязи X_1 и X_2 — разность
 $\mathbb{E}(X_1 | X_2 = 1) - \mathbb{E}(X_1 | X_2 = 0)$.

- › Корреляция Мэтьюса
- › Коэффициент V Крамера
- › Далее: оценка значимости корреляции

ЗНАЧИМОСТЬ КОРРЕЛЯЦИИ

ПРОДАЖИ МОРОЖЕНОГО

- » $r_{X_1 X_2} = 0.45, \rho_{X_1 X_2} = 0.44$
- » Связаны ли объём продаж мороженого и средняя дневная температура?



КРИТЕРИЙ СТЪЮДЕНТА

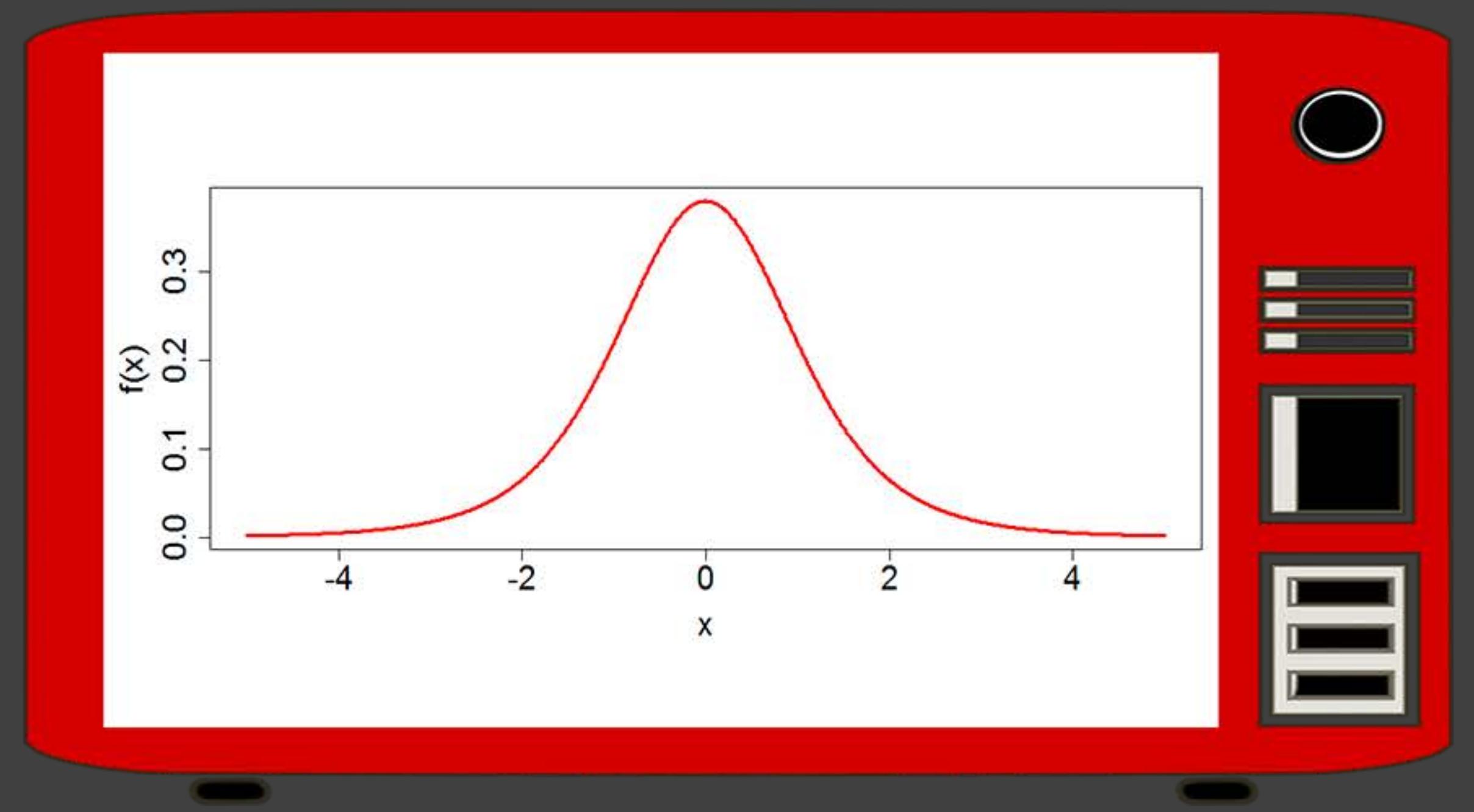
выборки: $(X_{1i}, X_{2i}), i = 1, \dots, n$

нулевая гипотеза: $H_0: r_{X_1 X_2} = 0$

альтернатива: $H_1: r_{X_1 X_2} < \neq > 0$

статистика: $T = \frac{r_{X_1 X_2} \sqrt{n - 2}}{\sqrt{1 - r_{X_1 X_2}^2}}$

нулевое распределение: $T \sim St(n - 2)$



ДЛЯ КОРРЕЛЯЦИИ СПИРМЕНА

- » Можно применять этот же критерий Стьюдента

ПРОДАЖИ МОРОЖЕНОГО

- » H_0 : линейной связи нет, $r_{X_1 X_2} = 0$
- » H_1 : признаки линейно связаны, $r_{X_1 X_2} \neq 0$
- » Критерий Стьюдента: $p = 3 \times 10^{-6}$, признаки линейно связаны, корреляция Пирсона — 0.45 (95% доверительный интервал — [0.28, 0.59]).

ПРОДАЖИ МОРОЖЕНОГО

- » H_0 : монотонной связи нет, $\rho_{X_1 X_2} = 0$
- » H_1 : признаки монотонно связаны, $\rho_{X_1 X_2} \neq 0$
- » Критерий Стьюдента: $p = 3 \times 10^{-6}$, признаки монотонно связаны, корреляция Спирмена — 0.44 (95% доверительный интервал — [0.26, 0.60]).

ЭФФЕКТИВНОСТЬ ЛЕЧЕНИЯ

- › Исследуется влияние сахарного диабета на эффективность тромболитической терапии.

	Выздоровели	Не выздоровели
Диабет	48	30
Нет	92	36

- › $MCC = -0.1074$.
- › Связано ли наличие диабета с эффективностью лечения?

КРИТЕРИЙ ХИ-КВАДРАТ

выборки: $(X_{1i}, X_{2i}), i = 1, \dots, n$

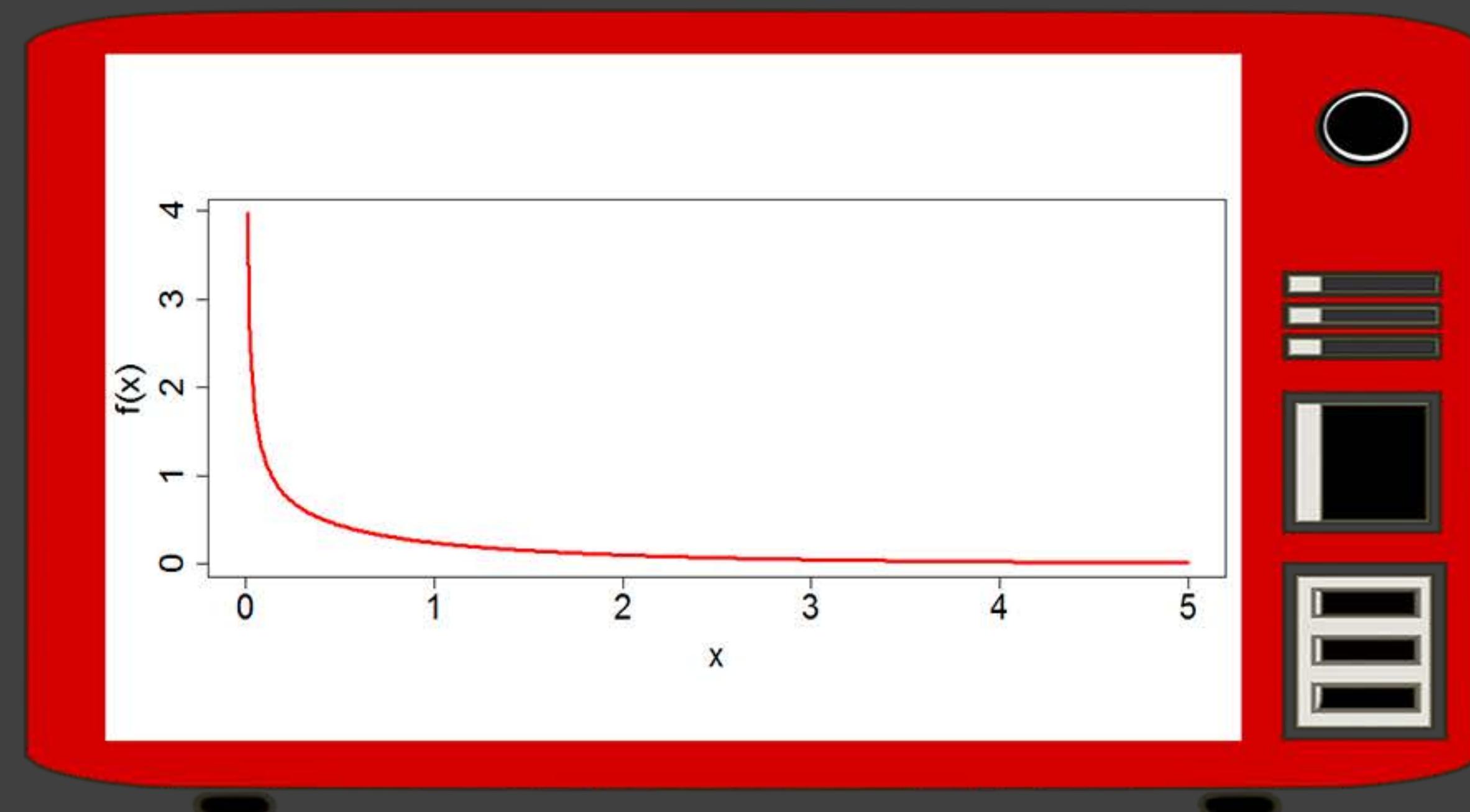
$X_1, X_2 \in \{0, 1\}$;

нулевая гипотеза: $H_0: \text{MCC}_{X_1 X_2} = 0$

альтернатива: $H_1: \text{MCC}_{X_1 X_2} \neq 0$

статистика: $\chi^2 = n \text{MCC}_{X_1 X_2}^2$

нулевое распределение: $\chi^2 \sim \chi_1^2$



» Условия применимости:

- ▶ $n \geq 40$
- ▶ $\frac{(a+c)(a+b)}{n}, \frac{(a+c)(c+d)}{n}, \frac{(b+d)(a+b)}{n}, \frac{(b+d)(c+d)}{n} > 5$

ЭФФЕКТИВНОСТЬ ЛЕЧЕНИЯ

	Выздоровели	Не выздоровели
Диабет	48	30
Нет	92	36

- › МСС = -0.1074
- › H_0 : эффективность лечения не зависит от наличия диабета
- › H_1 : эффективность лечения зависит от наличия диабета

ЭФФЕКТИВНОСТЬ ЛЕЧЕНИЯ

	Выздоровели	Не выздоровели
Диабет	48	30
Нет	92	36

- › МСС = -0.1074
- › Критерий хи-квадрат: $p = 0.1651$, нельзя утверждать, что связь есть.

КАТЕГОРИАЛЬНЫЕ ПРИЗНАКИ

X_1	X_2	1	\dots	j	\dots	K_2	\sum
1							
\vdots							
i				n_{ij}		n_{i+}	
\vdots							
K_1							
\sum				n_{+j}			n

КРИТЕРИЙ ХИ-КВАДРАТ

выборки: $(X_{1i}, X_{2i}), i = 1, \dots, n$

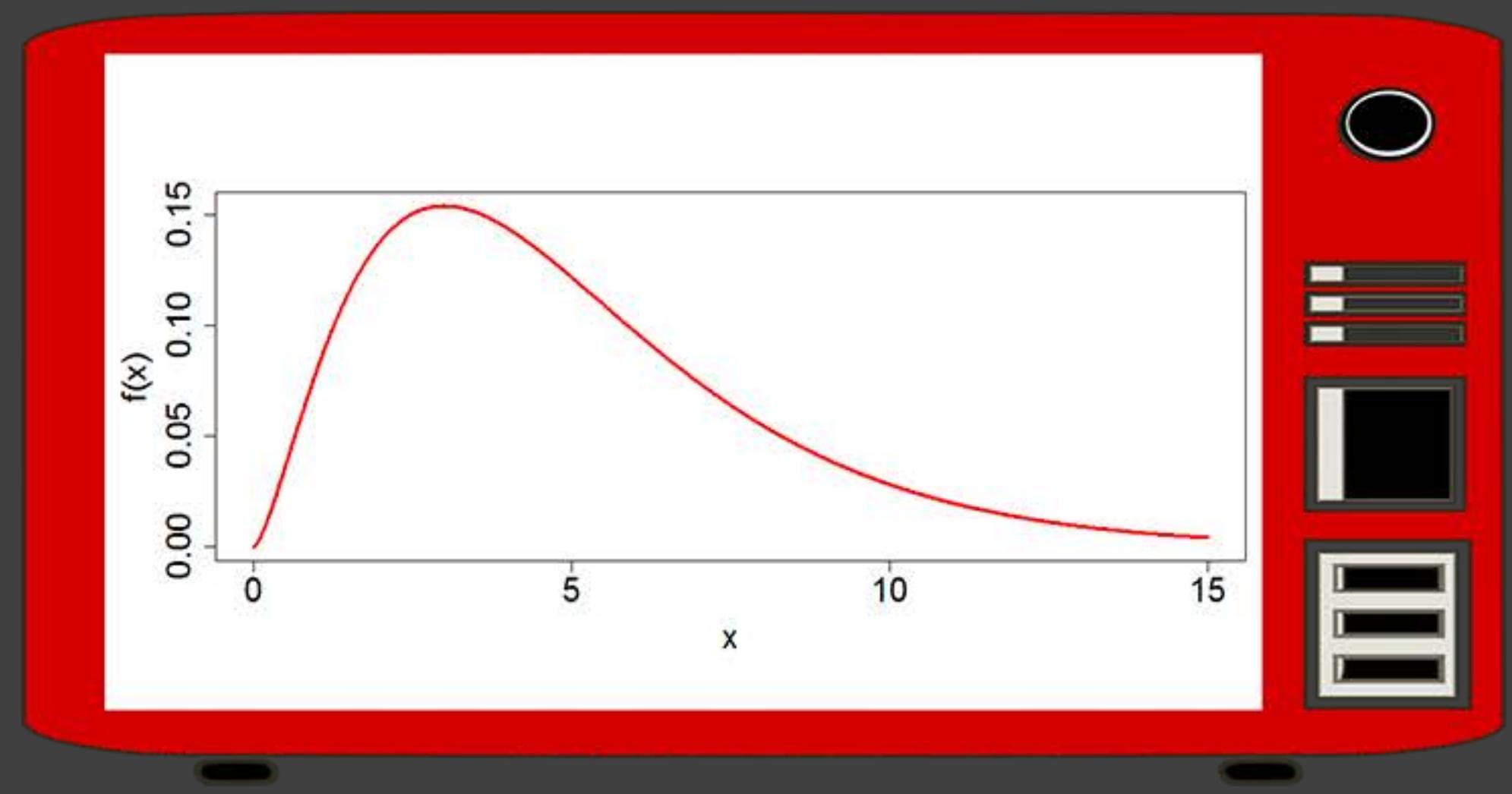
$X_1 \in \{1, \dots, K_1\}, X_2 \in \{1, \dots, K_2\}$

нулевая гипотеза: $H_0: X_1$ и X_2 независимы

альтернатива: $H_1: H_0$ неверна

статистика: $\chi^2(X_1^n, X_2^n) = \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} \frac{(n_{ij} - \frac{n_i + n_{+j}}{n})^2}{\frac{n_i + n_{+j}}{n}} =$
 $= n \left(\sum_{i=1}^{K_1} \sum_{j=1}^{K_2} \frac{n_{ij}^2}{n_{i+} n_{+j}} - 1 \right)$

$$\chi^2(X_1^n, X_2^n) \sim \chi^2_{(K_1-1)(K_2-1)}$$



КРИТЕРИЙ ХИ-КВАДРАТ

» Условия применимости:

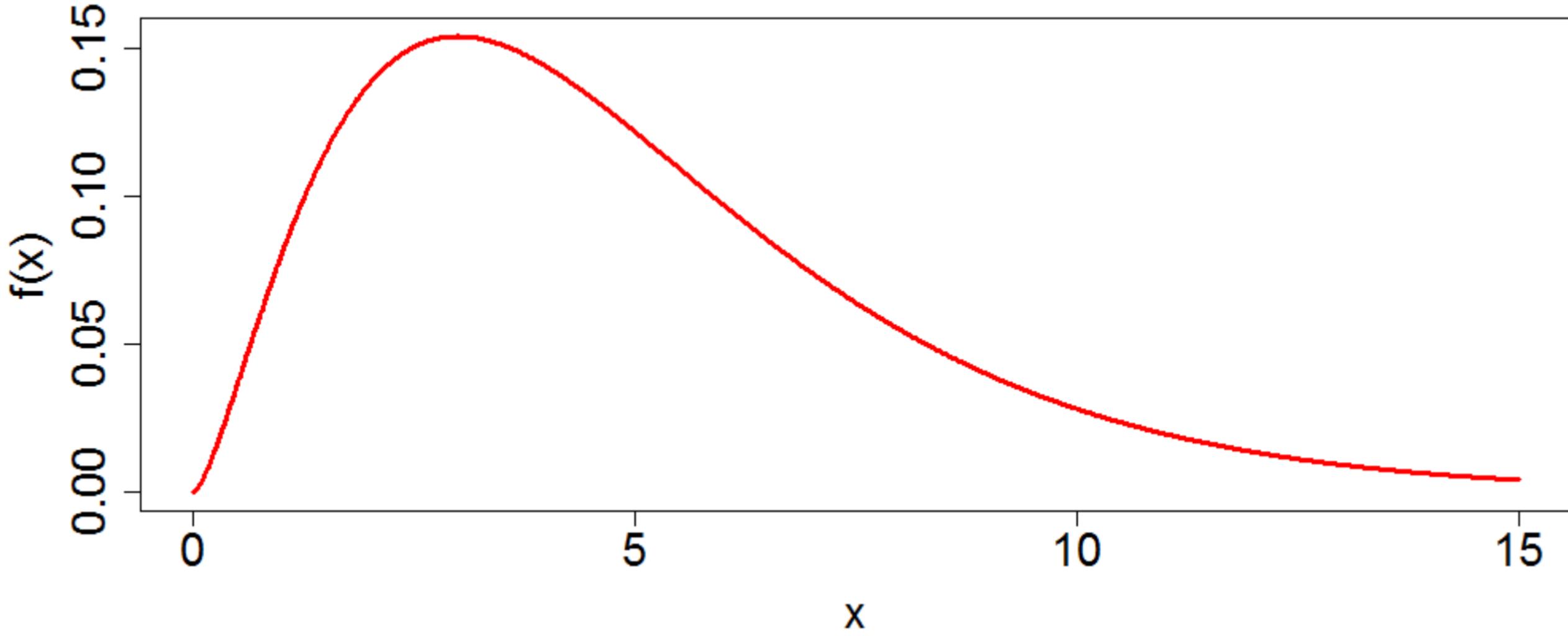
- ▶ $n \geq 40$
- ▶ $\frac{n_{i+j} + n_{+j}}{n} < 5$ не более, чем в 20% ячеек

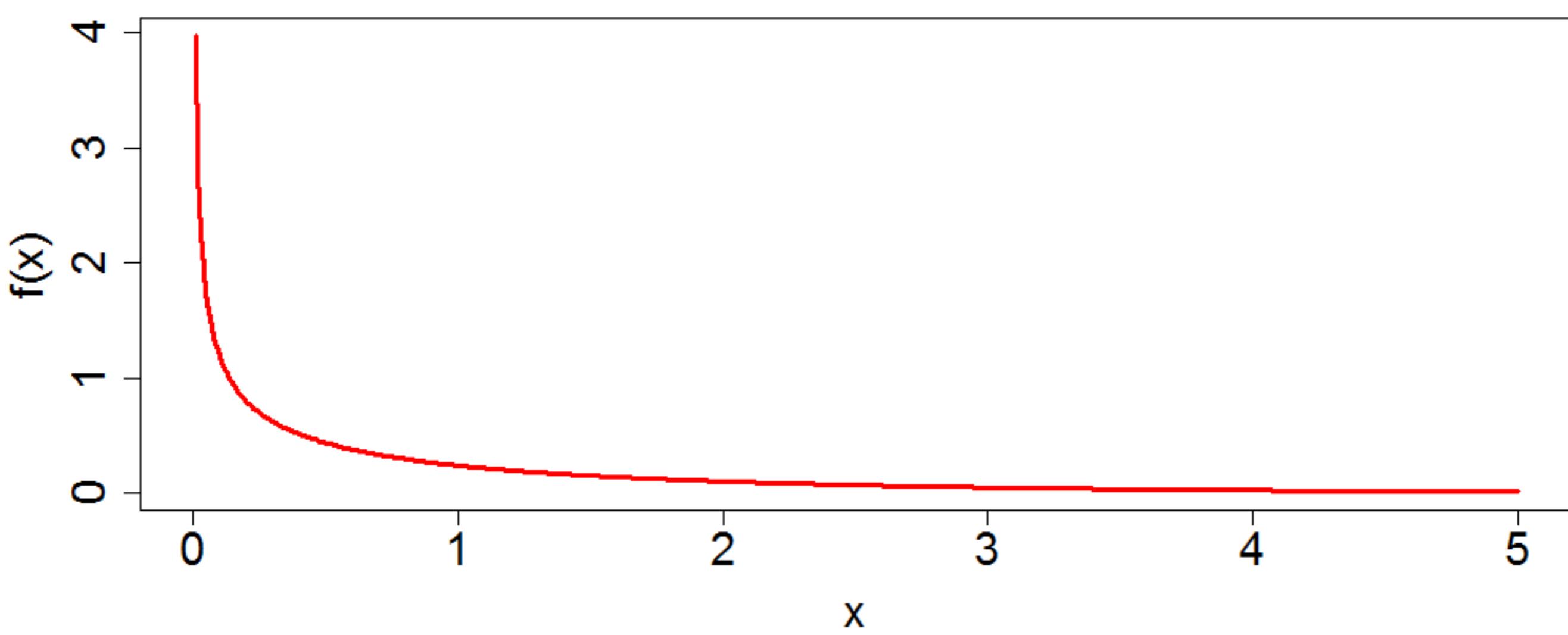
КРИТЕРИЙ ХИ-КВАДРАТ

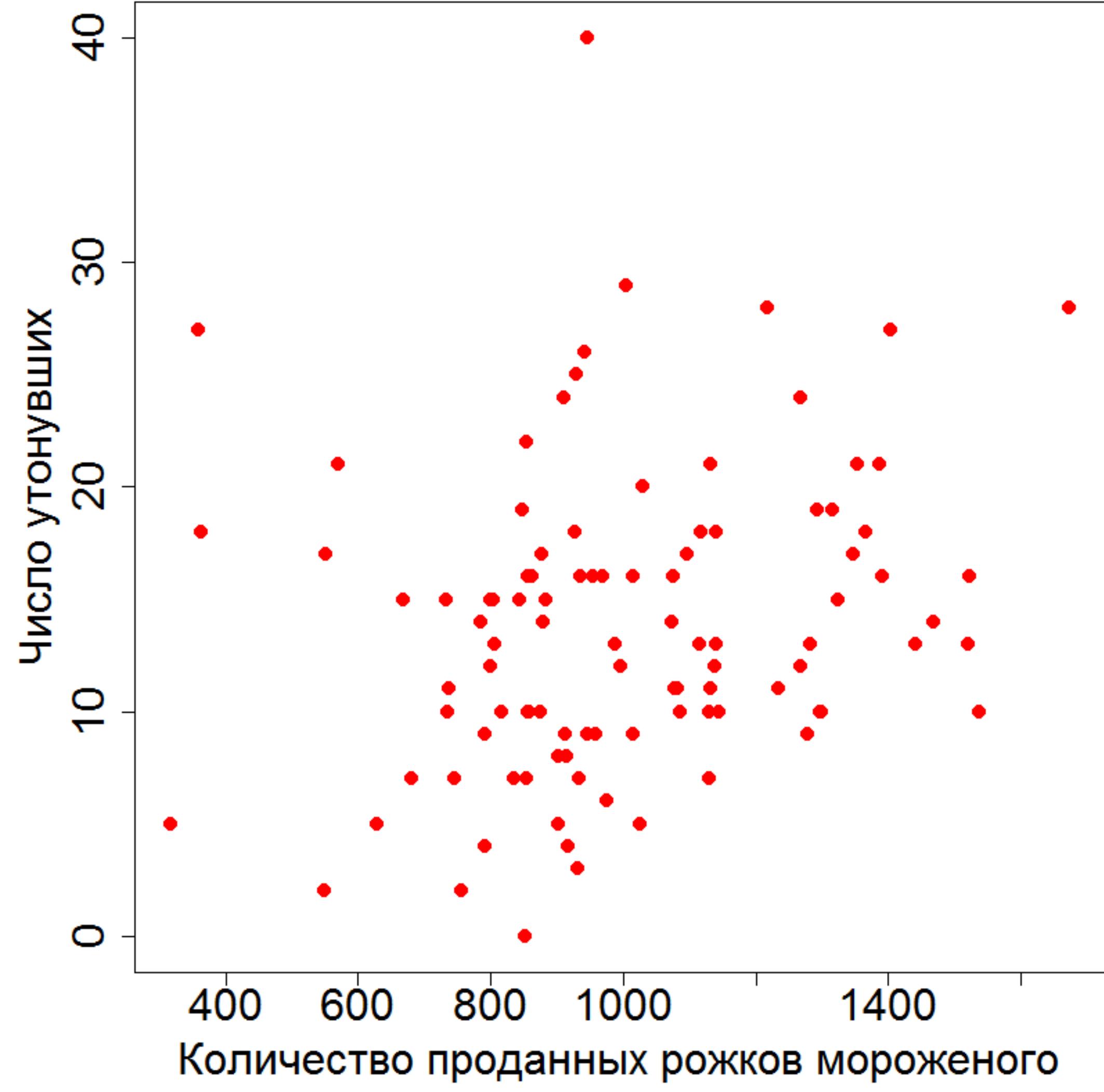
- » Проверяет значимость отличия от нуля коэффициента V Крамера:

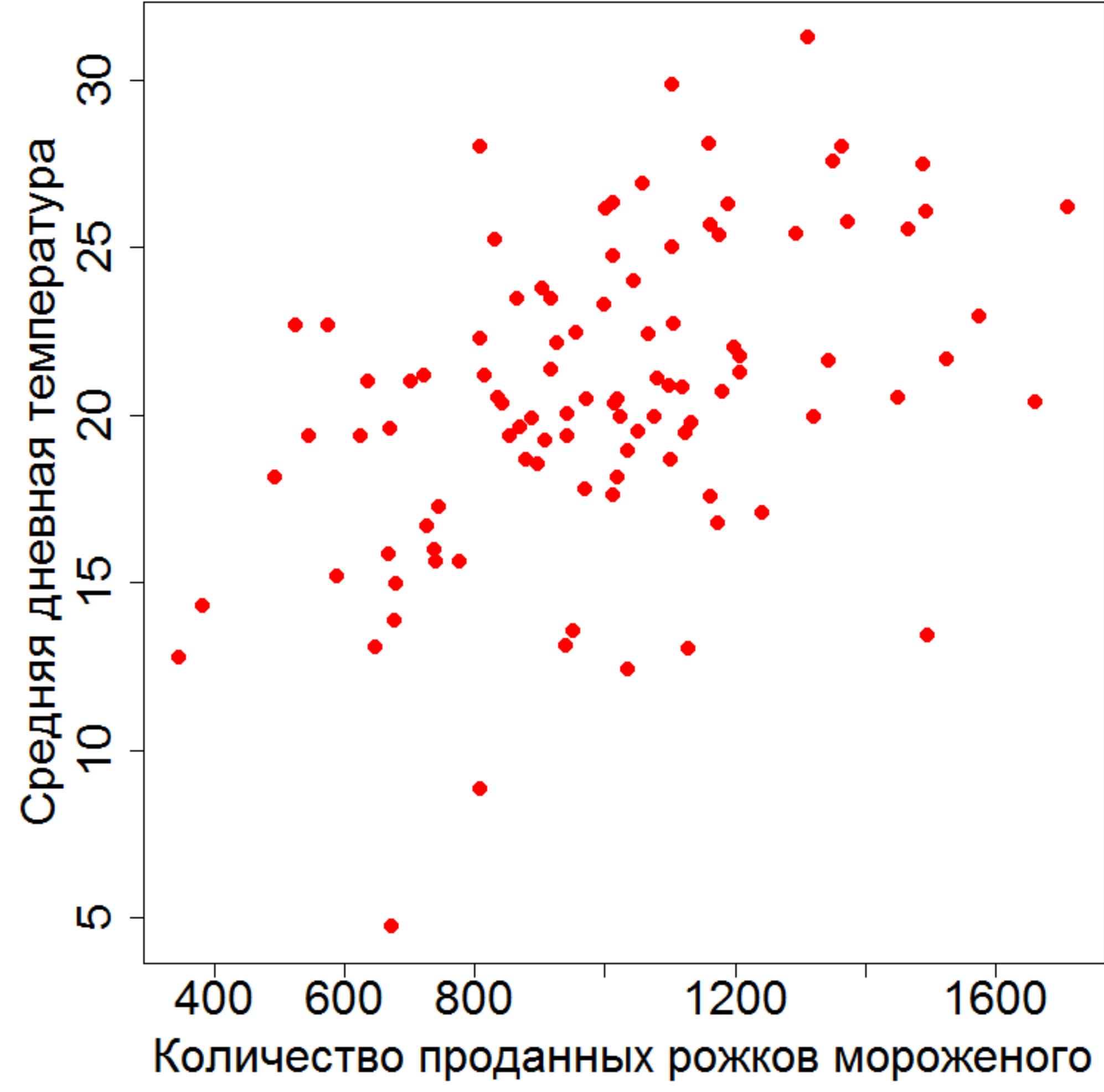
$$\phi_c(X_1^n, X_2^n) = \sqrt{\frac{\chi^2(X_1^n, X_2^n)}{n(\min(K_1, K_2) - 1)}}$$

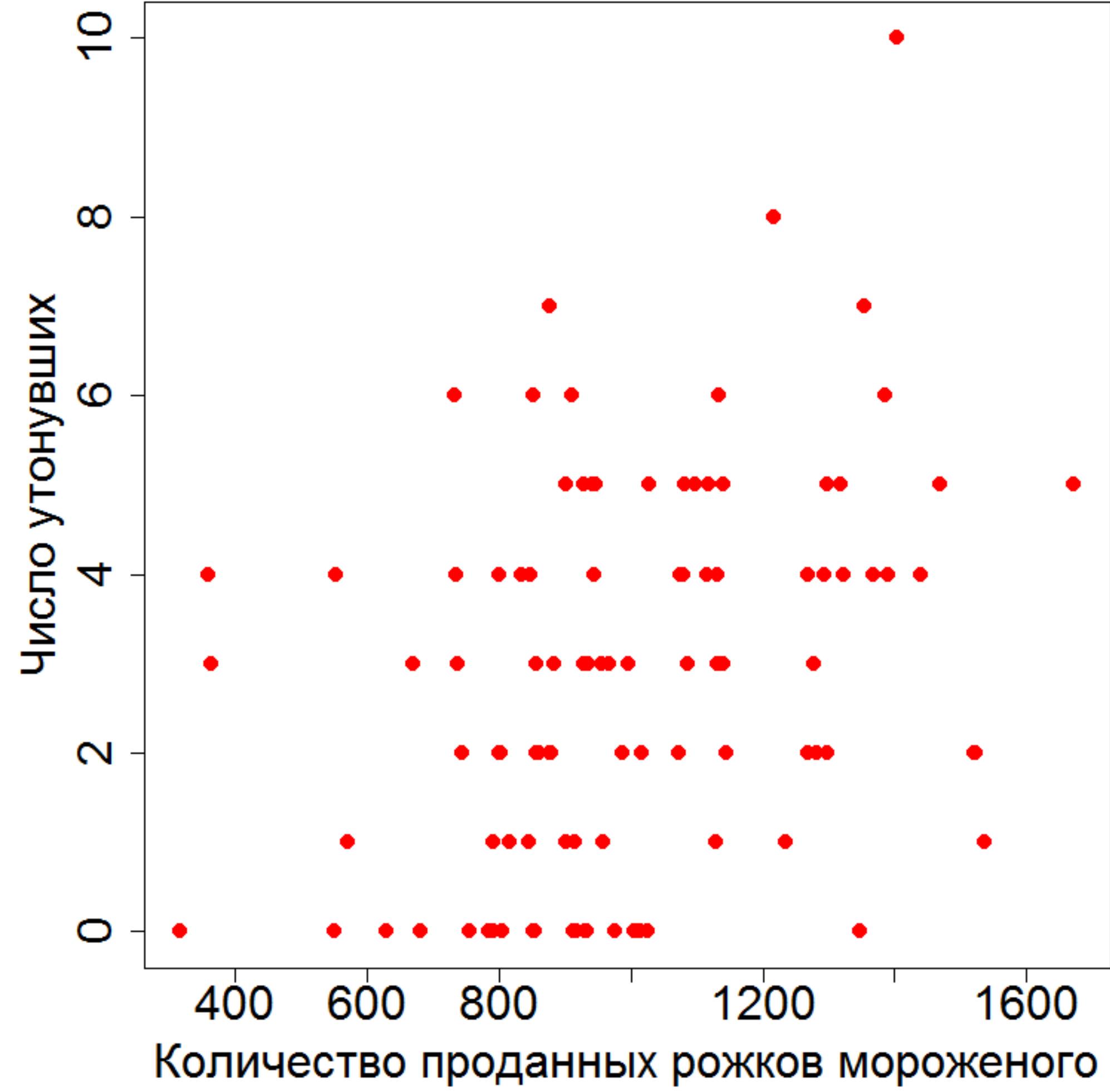
- › Критерии для проверки значимости корреляции:
 - ▶ критерий Стьюдента для корреляций Пирсона и Спирмена
 - ▶ критерий хи-квадрат для корреляции Мэтьюса и коэффициента V Крамера
- › Далее: интерпретация коэффициентов корреляции

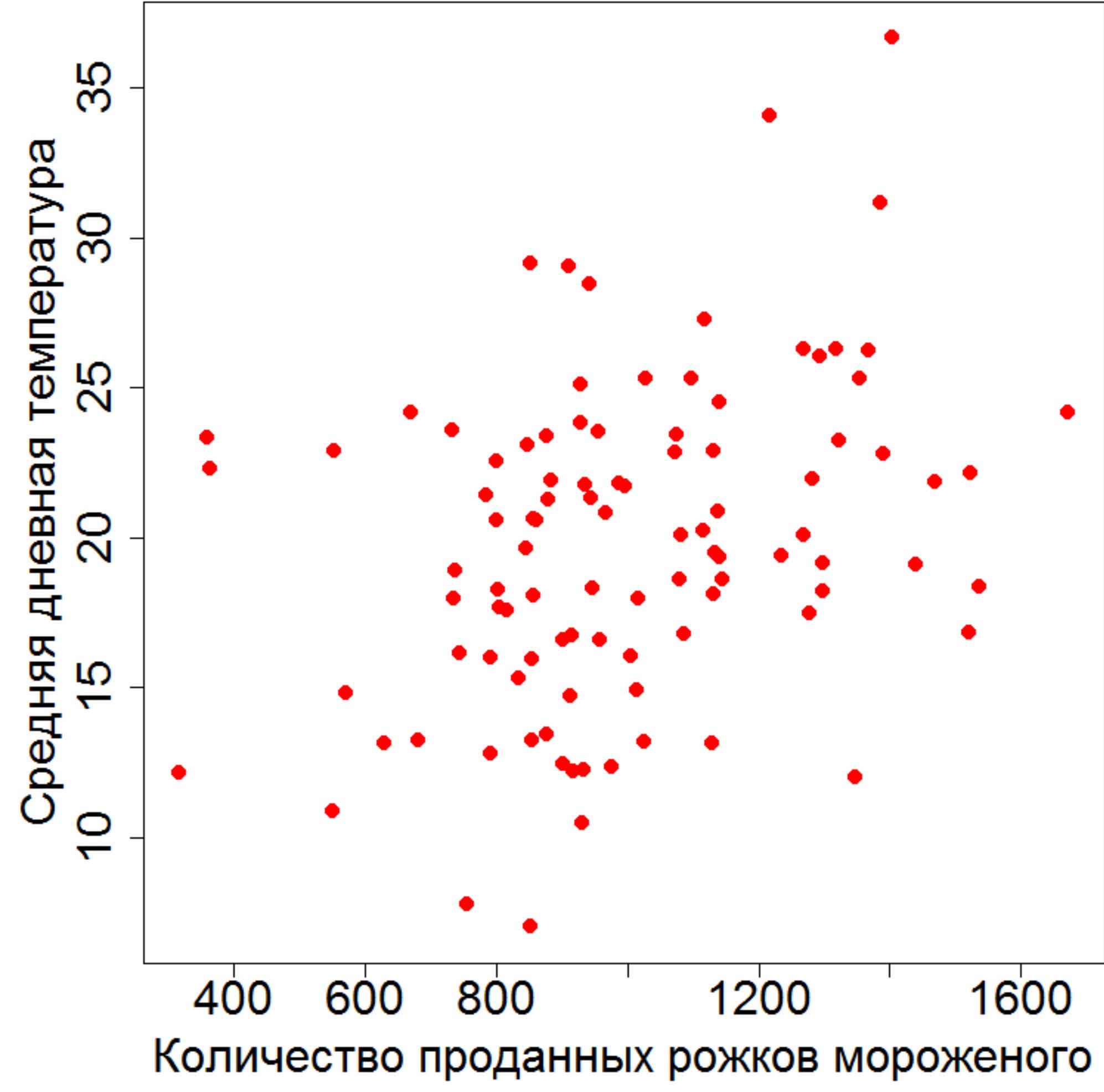


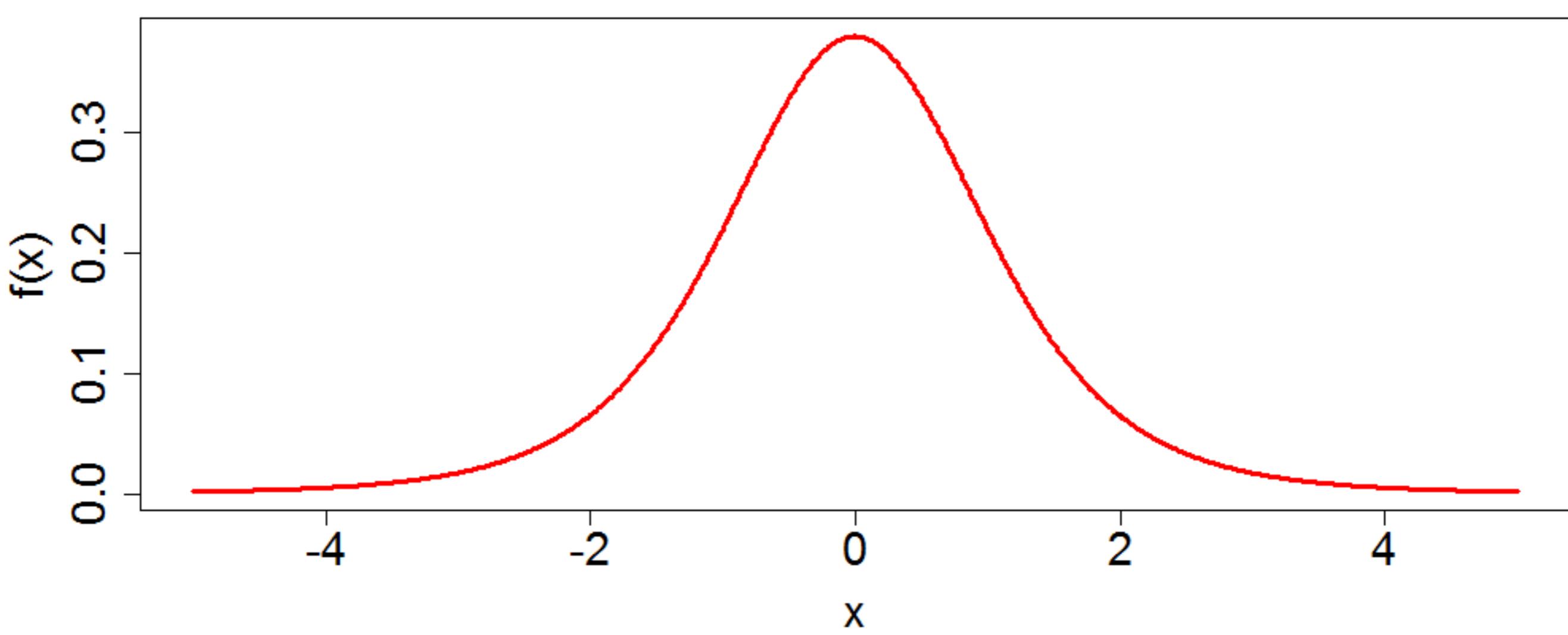












БУЛЩИТ И КОНСЕРВАТИВНОСТЬ

БУЛЩИТ И КОНСЕРВАТИВНОСТЬ



› <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0153419>

A red smartphone is shown from a side-on perspective, displaying a white web page. The phone's screen shows the PLOS ONE logo at the top left, followed by a yellow horizontal bar. Below the bar, the text 'RESEARCH ARTICLE' is visible. The main title of the article is 'Misperceiving Bullshit as Profound Is Associated with Favorable Views of Cruz, Rubio, Trump and Conservatism'. Below the title, the authors' names are listed: 'Stefan Pfattheicher¹*, Simon Schindler²'. At the bottom of the screen, there is a note: '1 Ulm University, Ulm, Germany, 2 Kassel University, Kassel, Germany'. The background of the slide is dark grey.



RESEARCH ARTICLE

Misperceiving Bullshit as Profound Is Associated with Favorable Views of Cruz, Rubio, Trump and Conservatism

Stefan Pfattheicher^{1*}, Simon Schindler²

1 Ulm University, Ulm, Germany, **2** Kassel University, Kassel, Germany



ОПРЕДЕЛЕНИЕ

- Буллит — бессодержательное, нелогичное или явно противоречащее элементарным научным знаниям утверждение
- Примеры:
 - ▶ “Скрытый смысл трансформирует беспрецедентную абстрактную красоту”

ОПРЕДЕЛЕНИЕ

- Буллит — бессодержательное, нелогичное или явно противоречащее элементарным научным знаниям утверждение
- Примеры:
 - ▶ “Воображение лежит в основе экспоненциальных пространственно-временных событий”

- Испытуемые: 196 граждан Америки (43.4% женщин, средний возраст 36.4 лет), набранные на Amazon Mechanical Turk
- Задание:
 - ▶ Оценить глубокомысленность утверждений по шкале от 1 ("абсолютно не глубокое") до 5 ("очень глубокое")

» Задание:

- ▶ Оценить степень симпатии к трём кандидатам в президенты США от демократической и трём — от республиканской партий по шкале от 1 (“очень несимпатичен”) до 5 (“очень симпатичен”)

» Задание:

- ▶ Оценить степень консервативности собственных политических взглядов по семибалльной шкале Лайкера.

- » Часть утверждений, оцениваемых респондентами, — буллит, часть — относительно редкие поговорки (“Промокший человек не боится дождя”)

- » Для каждого испытуемого вычислялась средняя склонность считать буллит глубокомысленным
- » Вычислялась корреляция Спирмена между итоговым признаком, консервативностью политических взглядов и степенями симпатии к кандидатам

- Вычислялась корреляция Спирмена между итоговым признаком, консервативностью политических взглядов и степенями симпатии к кандидатам

- Для проверки значимости отличия корреляции от нуля использовался критерий Стьюдента.

РЕЗУЛЬТАТ

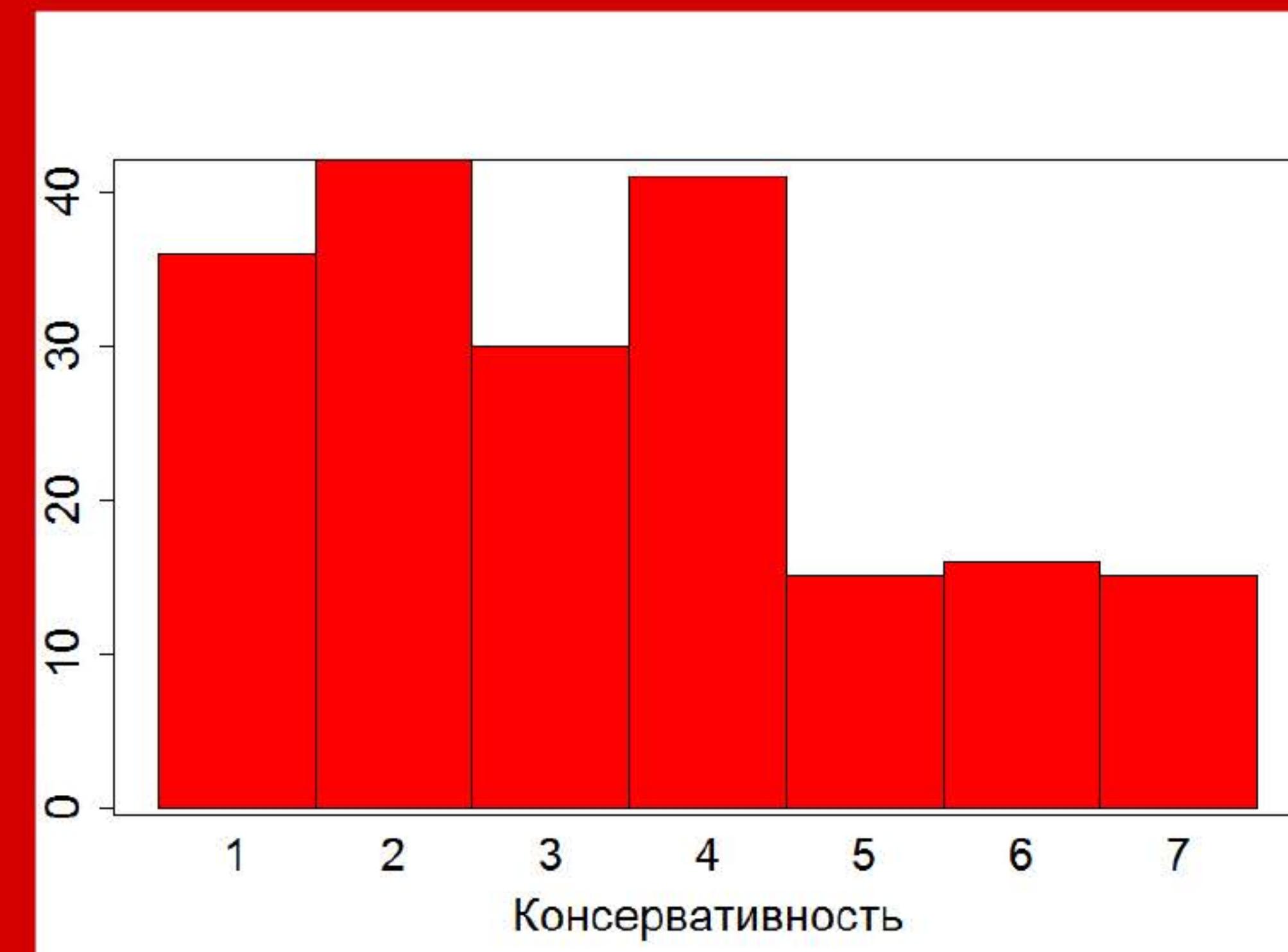


- Обнаружена значимая положительная корреляция между тягой к буллшу и:
 - ▶ Симпатией к Теду Крузу, Марку Рубио и Дональду Трампу
 - ▶ Степенью консерватизма

- » Репрезентативность выборки: аудитория Amazon Mechanical Turk — не случайная выборка граждан США

ПРОБЛЕМЫ

- › Репрезентативность выборки
- › Несбалансированность выборки:

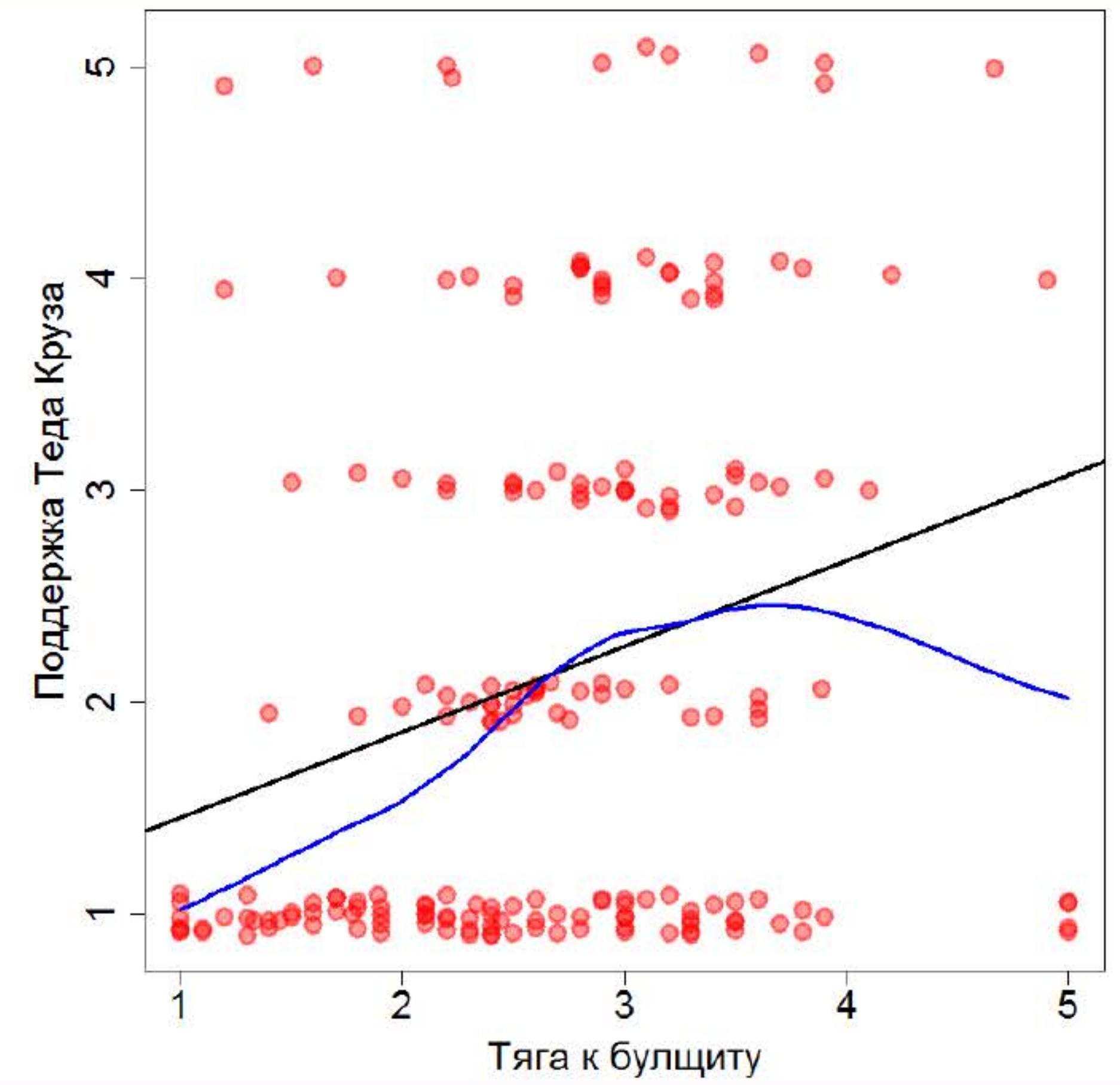


- › Репрезентативность выборки
- › Несбалансированность выборки
- › Сырые данные

СЫРЫЕ ДАННЫЕ



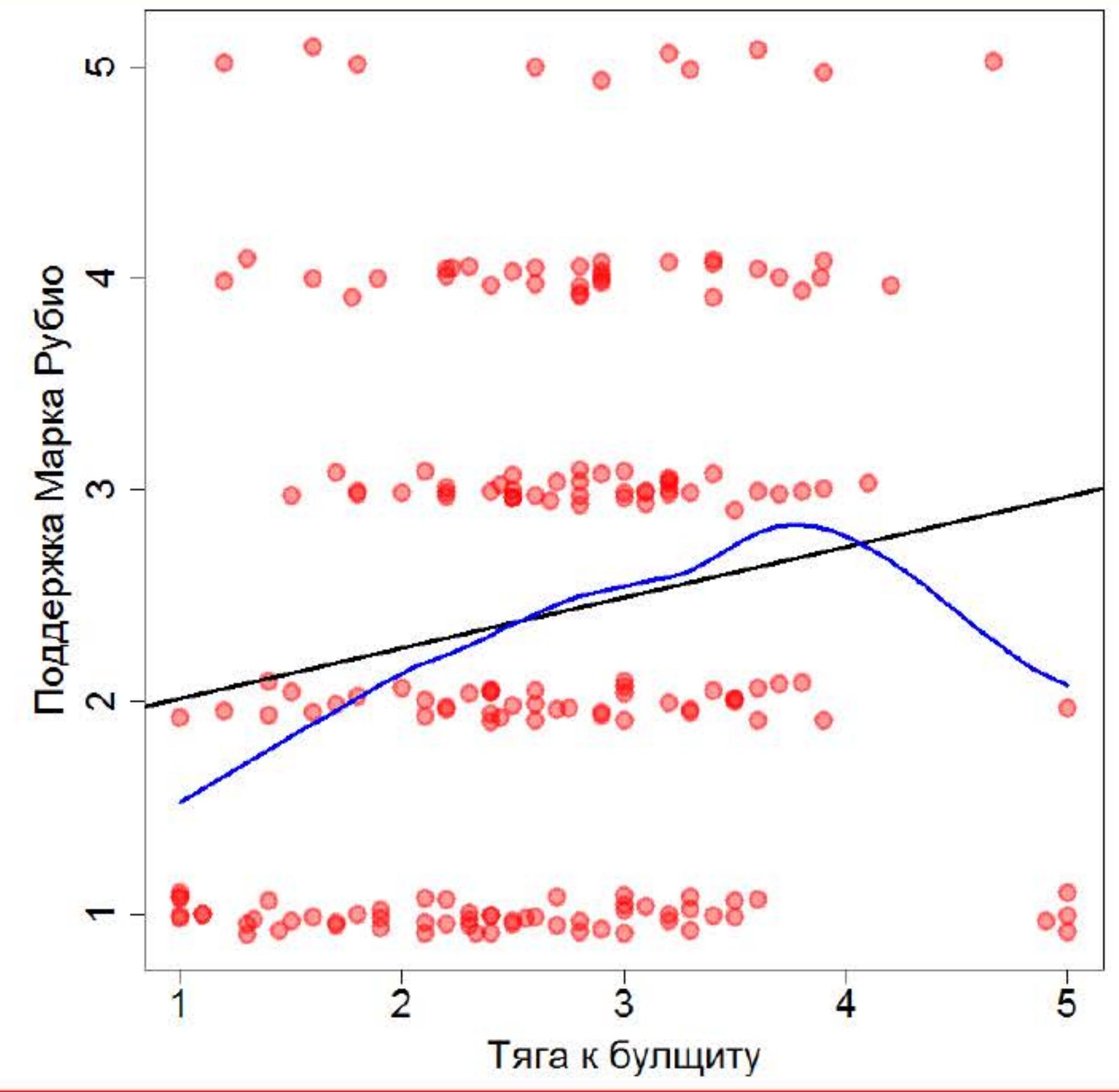
» $\rho_{XY} = 0.3, \ p = 2 \times 10^{-5}$



СЫРЫЕ ДАННЫЕ



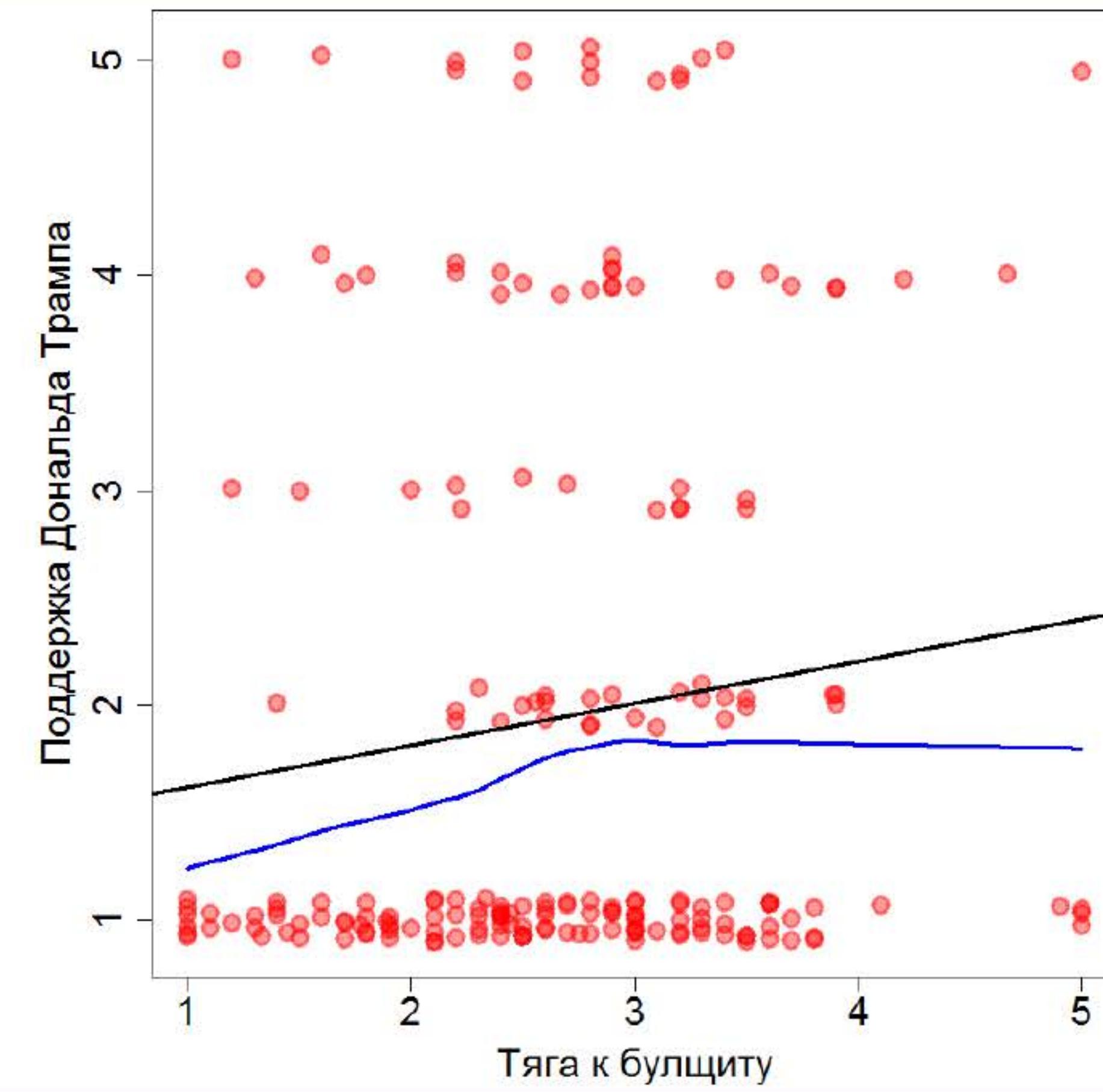
» $\rho_{XY} = 0.2, \ p = 0.0064$



СЫРЫЕ ДАННЫЕ



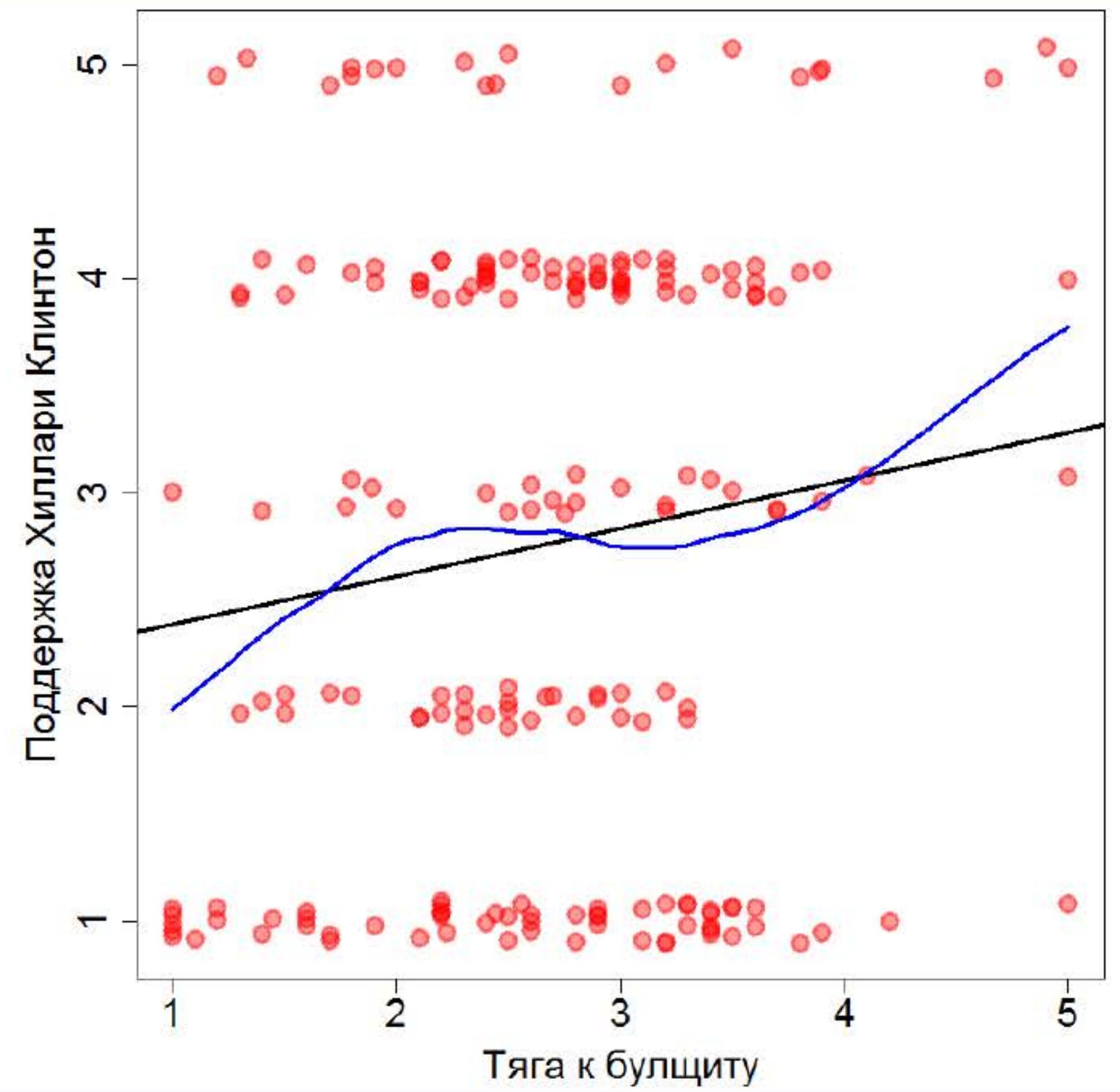
» $\rho_{XY} = 0.15, \ p = 0.0324$

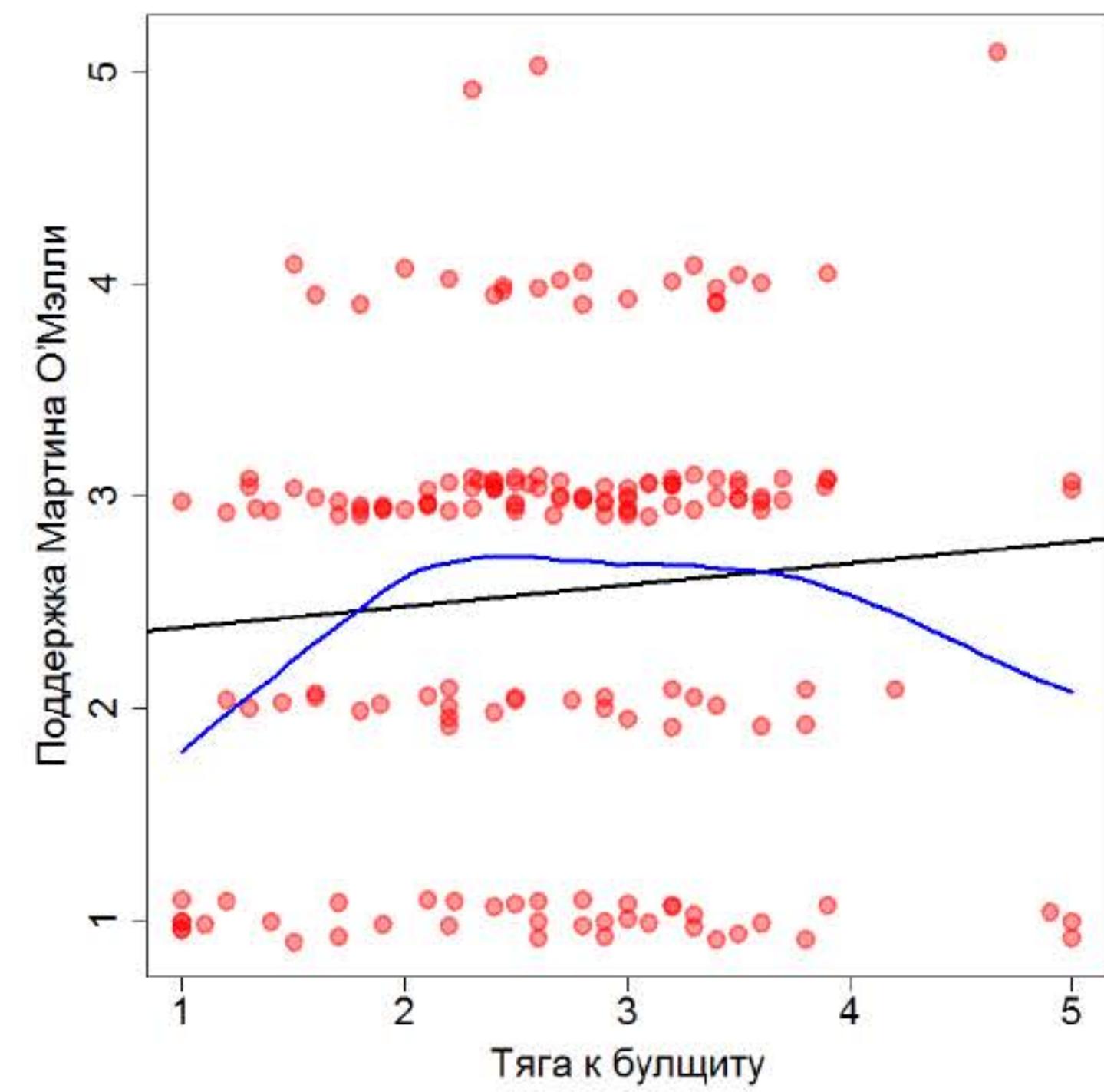
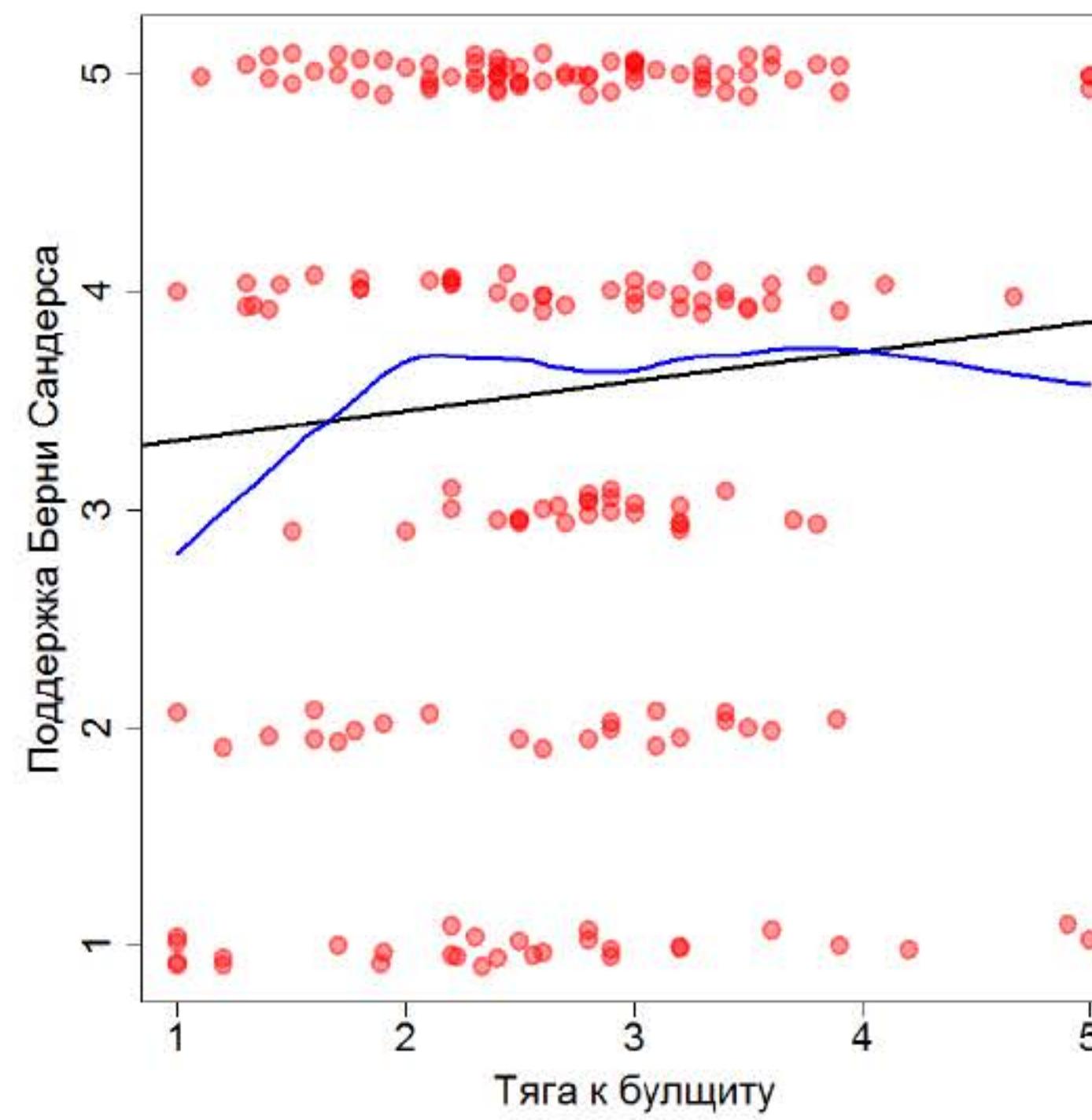
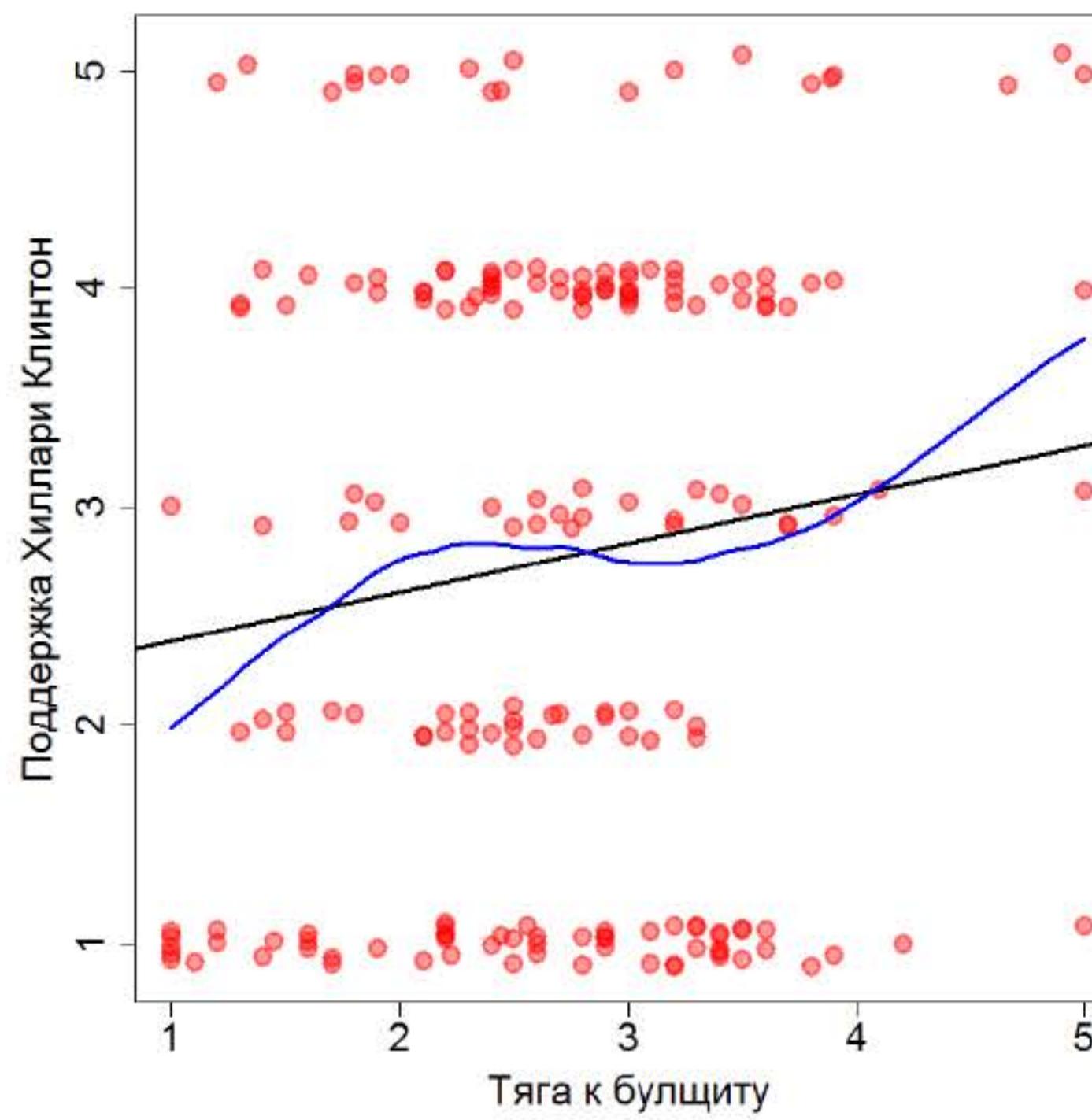
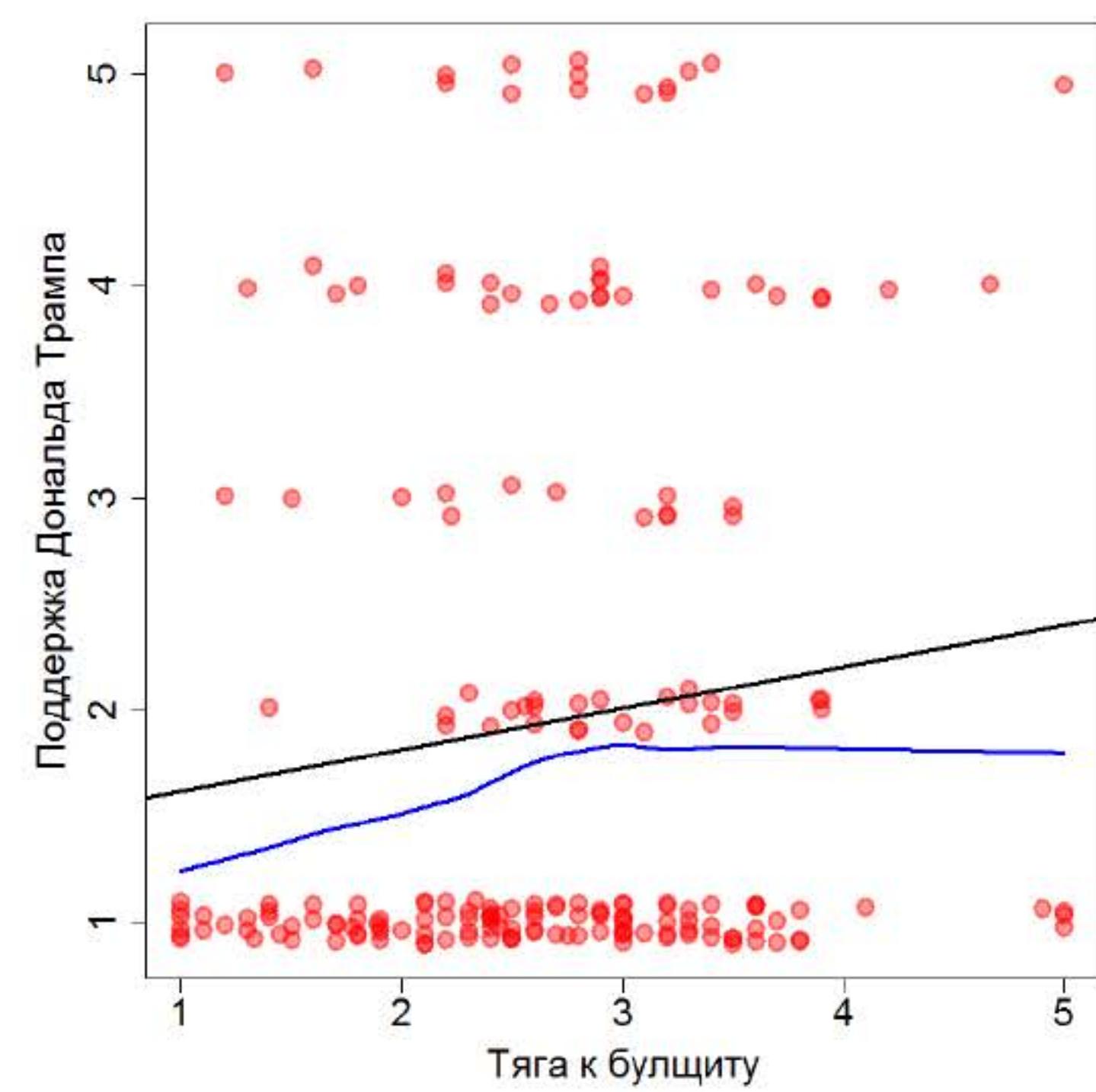
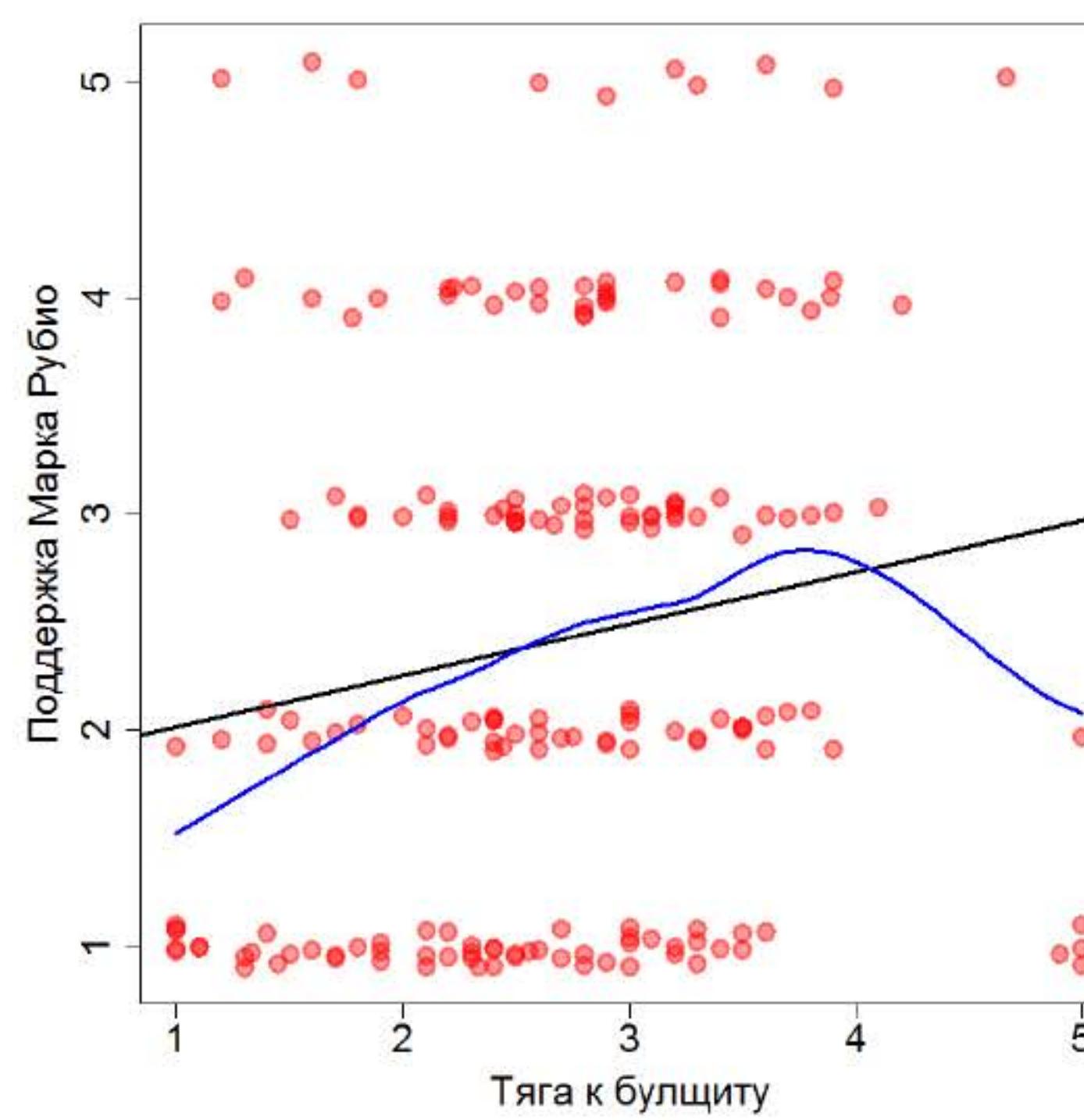
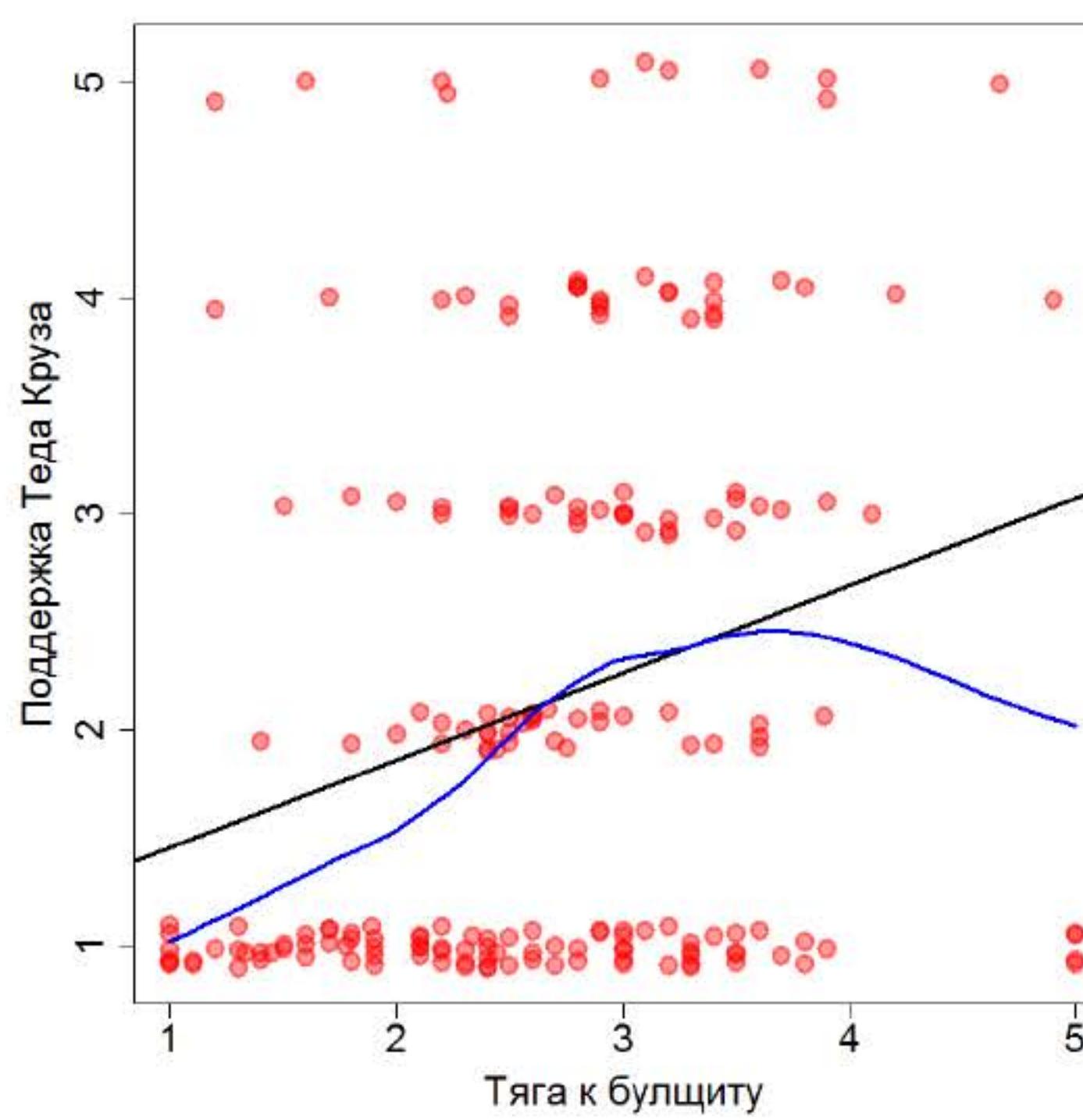


СЫРЫЕ ДАННЫЕ



» $\rho_{XY} = 0.09, \ p = 0.212$



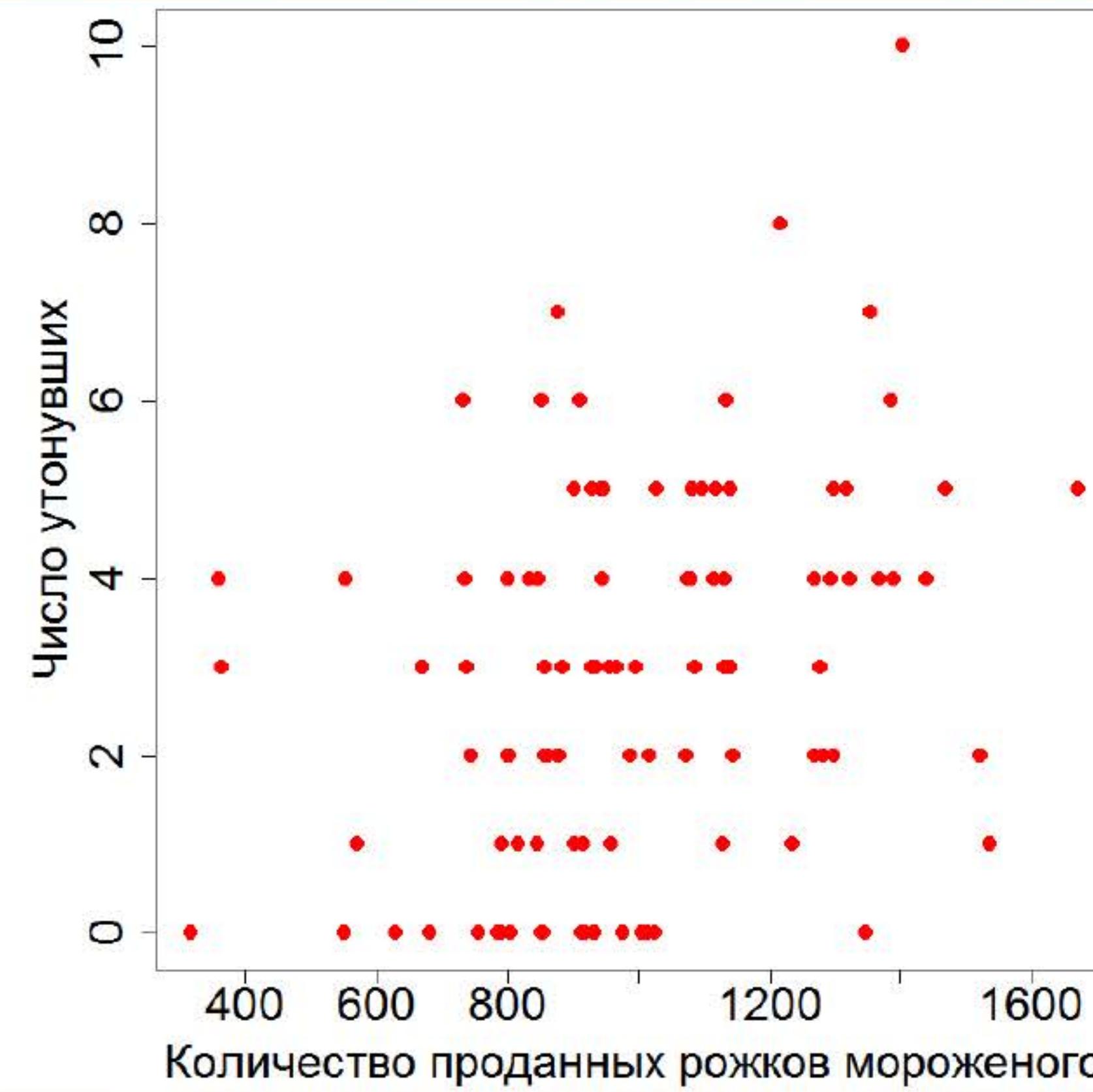


- › Всегда смотрите на сырье данные!
- › Корреляционный анализ — не сосисочная машина
(справедливо для всех статистических методов)
- › Далее: интерпретация коэффициентов корреляции

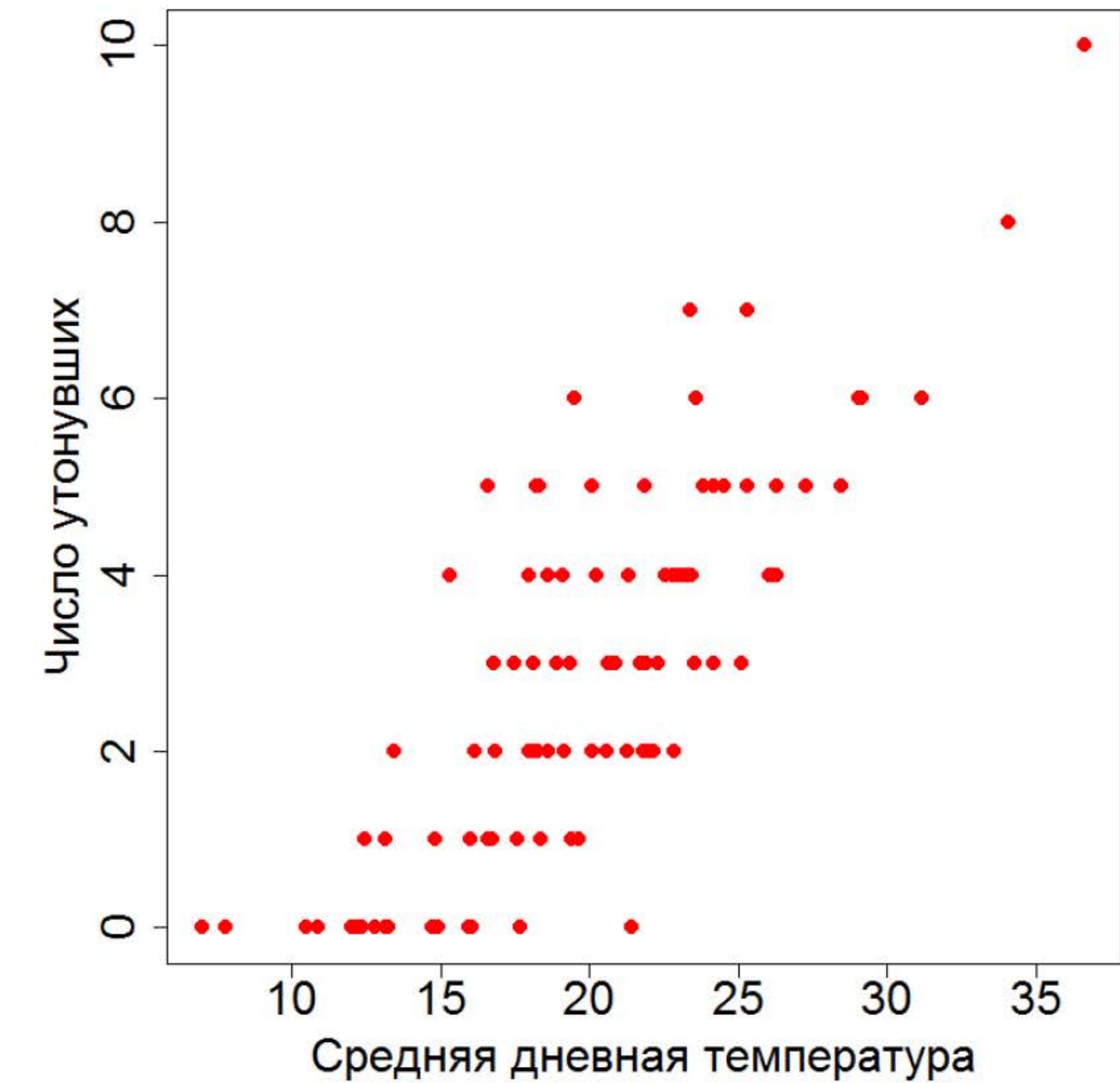
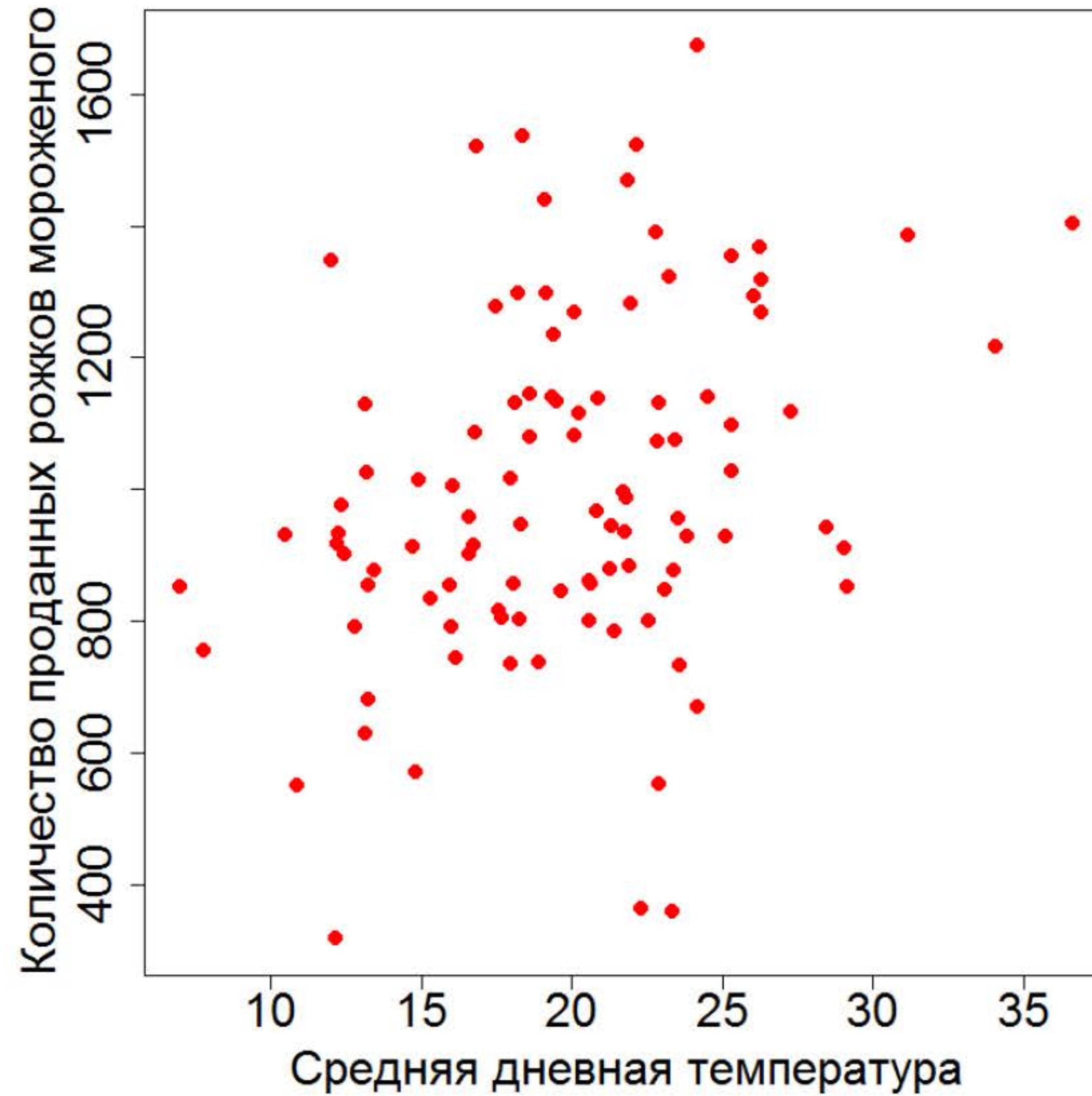
КОРРЕЛЯЦИЯ И ПРИЧИННО-СЛЕДСТВЕННАЯ СВЯЗЬ

МОРОЖЕНОЕ И СМЕРТЬ

- » $r_{X_1 X_2} = 0.33, p = 0.0009$ (критерий Стьюдента),
95% доверительный интервал — [0.138, 0.491]



МОРОЖЕНОЕ, СМЕРТЬ И ТЕМПЕРАТУРА



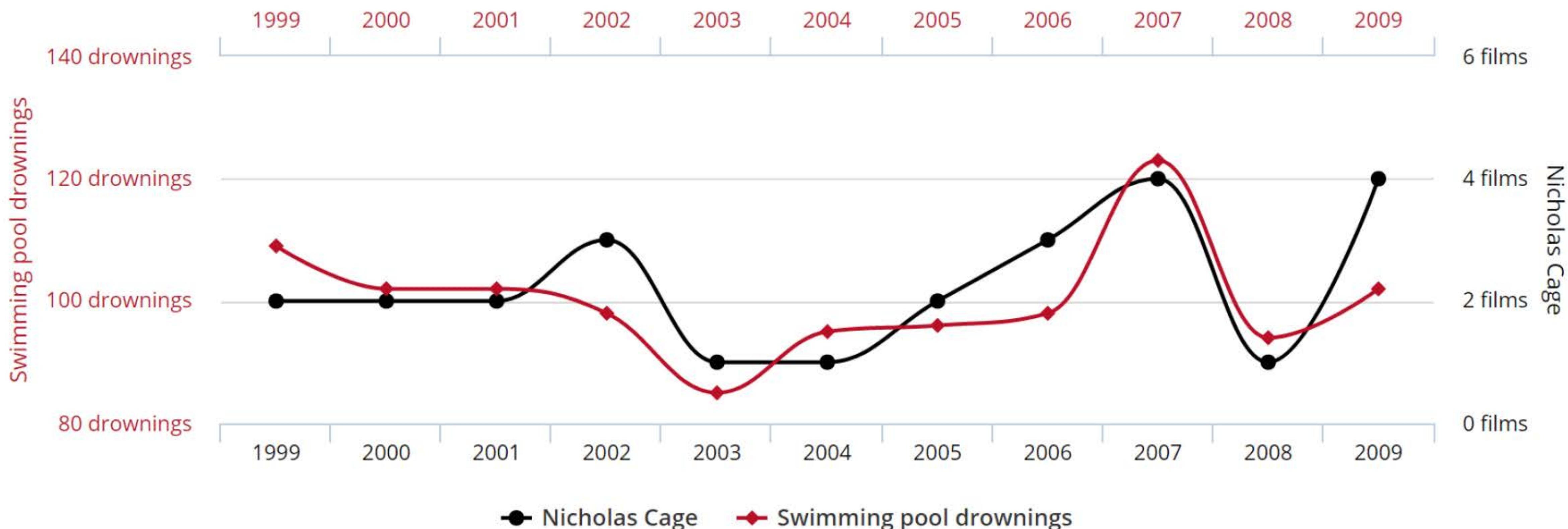
ДРУГИЕ ПРИМЕРЫ

- › Количество самоубийств и радиоприёмников на душу населения (размер города)
- › Уровень CO₂ в атмосфере и распространённость ожирения (уровень жизни)
- › Рыночная доля Internet Explorer и количество убийств в США (время)

СМЕРТЬ И НИКОЛАС КЕЙДЖ

- $r_{X_1 X_2} = 0.67, p = 0.0253$ (критерий Стьюдента),
95% доверительный интервал — [0.110, 0.905]

- Больше: <http://www.tylervigen.com>



- » Корреляция $\not\Rightarrow$ причинно-следственная связь
- » Причинно-следственная связь $\xrightarrow{\text{возможно}}$ корреляция
- » Далее: ассоциативные правила